

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 23, 2003, 11:16:55; Search time 1066.61 Seconds
(without alignments)
4101.602 Million cell updates/sec

Title: US-09-941-309-1

Perfect score: 180
Sequence: 1 ggaagctcccttcctcaag.....tcataagaatggaatgc 180

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 1215238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
EST:
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estnu:*
5: em_estov:*
6: em_estpi:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrc:*
21: em_gss_vrc:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	180	100.0	334	B0636767
2	180	100.0	335	B0636767
3	180	100.0	537	B0636767
4	180	100.0	553	B0636767

Result No.	Score	Query Match length	ID	Description
5	180	100.0	1115	B0636767
6	167.4	93.0	321	B0636767
7	165.2	91.8	220	B0636767
8	130.4	72.4	293	B0636767
9	130.4	72.4	325	B0636767
10	127.4	70.8	433	B0636767
11	120.8	67.1	285	B0636767
12	120.8	67.1	313	B0636767
13	120.8	67.1	318	B0636767
14	120.8	67.1	330	B0636767
15	120.8	67.1	332	B0636767
16	120.8	67.1	361	B0636767
17	120.8	67.1	420	B0636767
18	119.8	66.6	364	B0636767
19	119.4	66.3	316	B0636767
20	119.2	66.2	277	B0636767
21	117.6	65.3	395	B0636767
22	116.2	64.6	300	B0636767
23	114.4	63.6	277	B0636767
24	114.4	63.6	287	B0636767
25	112.8	62.7	300	B0636767
26	112.6	62.6	220	B0636767
27	112.6	62.6	229	B0636767
28	112	62.2	390	B0636767
29	111.2	61.8	277	B0636767
30	109.6	60.9	370	B0636767
31	109.2	60.7	369	B0636767
32	109	60.6	232	B0636767
33	108.2	60.1	350	B0636767
34	103.4	57.4	266	B0636767
35	100.8	56.0	242	B0636767
36	94	52.2	268	B0636767
37	90.4	50.2	282	B0636767
38	90.4	50.2	286	B0636767
39	88	48.9	214	B0636767
40	83.6	46.4	301	B0636767
41	83.4	46.3	146	B0636767
42	82	45.6	210	B0636767
43	81.2	45.1	835	B0636767
44	80.6	44.8	229	B0636767
45	69.8	38.8	402	B0636767

ALIGNMENTS

RESULT 1
B0636767
LOCUS
DEFINITION
B0636767
ACCESSION
B0636767
VERSION
B0636767.1
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS
Wistow G., Bernstein S.L., Wyatt M.K., Ray S., Behal A., Touchman J.W., Bouffard G., Smith D. and Peterson K.
TITLE
Expressed sequence tag analysis of human retina for the NEIRANK Project: Retinidin, an abundant, novel retinal cDNA and alternative splicing of other retina-preferred gene transcripts
Mol. Vis. 8 (4), 2002
JOURNAL
Contact: Wistow G.
COMMENT
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: graeme@helix.nih.gov
Plate: 01 row: b column: 10

FEATURES
source
Seq primer: M13RPI reverse primer (ABI).
Location/Qualifiers
1. .334
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="he01b10"
/tissue_type="Retina"
/dev_stage="Adult"
/lab_host="EMDHI08"
/clone_lib="Human Retina cdna (Un-normalized, unamplified)
): hd/he"
/note="Organ: Eye; Vector: pSPORI1; Neural retina tissue was dissected from two 80 year old donors with no observed eye disease. 100ug of total RNA was used for library construction. A directionally cloned cDNA library in the pSPORI1 vector (Life Technologies) was constructed at Bioserve Biotechnology (Laurel MD) essentially following details of the SuperScript Plasmid System full instruction manual (http://www.lifetech.com/). First strand synthesis was carried out using a Not I primer-adaptor [5'-pGACTAGTCTAGATCGGACGGCGCC(T)15-3']. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

BASE COUNT
ORIGIN
94 a 86 c 88 g 66 t

Query Match
Best Local Similarity 100.0%; Score 180; DB 13; Length 334;
Pred. No. 5.1e-45;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAAGCTCCCTTTCAGAGATGCCATGTGAACACATGTAGATCTCAACCTGT 60
|||||
Db 63 GGAAGCTCCCTTTCAGAGATGCCATGTGAACACATGTAGATCTCAACCTGT 122
|||||
QY 61 TCCAGATGTCCAACTGGTCTGGGACATGATGGCTCATATATAGATGAATGCCAG 120
|||||
Db 123 TCCAGATGTCCAACTGGTCTGGGACATGATGGCTCATATATAGATGAATGCCAG 182
|||||
QY 121 CTCTGCTTGGCCCGGATAAACCAACAGACATCCAGATCATGAAGAATGCCAATGC 180
|||||
Db 183 CTCTGCTTGGCCCGGATAAACCAACAGACATCCAGATCATGAAGAATGCCAATGC 242
|||||

RESULT 2
BM856263 355 bp mRNA linear EST 06-MAR-2002
LOCUS K-EST0139933 S14K402 Homo sapiens CDNA clone S14K402-51-E10 5',
DEFINITION mRNA sequence.
ACCESSION BM856263
VERSION BM856263.1 GI:19212662
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 355)
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
21C Frontier Korean EST Project 2001
TITLE Unpublished
JOURNAL
COMMENT
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 51 row: E column: 10
High quality sequence stop: 355.
Location/Qualifiers

source
1. .355
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="S14K402-51-E10"
/cell_line="K402"
/lab_host="Top10F"
/clone_lib="S14K402"
/note="Organ: Stomach; Vector: pTZ19RPI; Site:1: EcorI;
Site:2: NotI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then deapped
with tabacco acid pyrophosphatase (TAP). The deapped
intact mRNA was ligated with DNA-RNA linker including Ecor
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcorI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."

BASE COUNT
ORIGIN
90 a 102 c 94 g 69 t

Query Match
Best Local Similarity 100.0%; Score 180; DB 12; Length 355;
Pred. No. 5.2e-45;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAAGCTCCCTTTCAGAGATGCCATGTGAACACATGTAGATCTCAACCTGT 60
|||||
Db 119 GGAAGCTCCCTTTCAGAGATGCCATGTGAACACATGTAGATCTCAACCTGT 178
|||||
QY 61 TCCAGATGTCCAACTGGTCTGGGACATGATGGCTCATATATAGATGAATGCCAG 120
|||||
Db 179 TCCAGATGTCCAACTGGTCTGGGACATGATGGCTCATATATAGATGAATGCCAG 238
|||||
QY 121 CTCTGCTTGGCCCGGATAAACCAACAGACATCCAGATCATGAAGAATGCCAATGC 180
|||||
Db 239 CTCTGCTTGGCCCGGATAAACCAACAGACATCCAGATCATGAAGAATGCCAATGC 298
|||||

RESULT 3
BG207464 537 bp mRNA linear EST 21-APR-2001
LOCUS RST26945 Athersys RAGE Library Homo sapiens CDNA, mRNA sequence.
DEFINITION BG207464
ACCESSION BG207464
VERSION BG207464.1 GI:13729151
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 537)
Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.,
Lerner,L., Costanzo,D., McElligott,K., Boozar,S., Mays,R., Smith
,E., Veloso,N., Kikka,A., Hess,J., Cottrien,K., Lo,K., Offenbacher
,J., Danzig,J. and Ducar,M.
Creation of genome-wide protein expression libraries using random
activation of gene expression
Nat. Biotechnol. 19 (5), 440-445 (2001)
TITLE
JOURNAL
MEDLINE 21227151
PUBMED 11329013
COMMENT
Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@athersys.com
High quality sequence stop: 482.

FEATURES
source
Location/Qualifiers
1. .537
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_id="HT1080"
/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, In press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

BASE COUNT
150 a 133 c 148 g 106 t

ORIGIN

Query Match 100.0%; Score 180; DB 10; Length 537;
Best Local Similarity 100.0%; Pred. No. 6e-45;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 GGAAGCTCCCTTTCTCAAGAAATGCCATCTGTGAACAGATGTAAGTCTCCAACTGT 60
|||||
270 GGAAGCTCCCTTTCTCAAGAAATGCCATCTGTGAACAGATGTAAGTCTCCAACTGT 329
|||||

Qy 61 TCCAGATGTCACACCTGCTGCGGCACTGATGGGCTCACATATACGAATGATGCCAG 120
|||||
Db 330 TCCAGATGTCACACCTGCTGCGGCACTGATGGGCTCACATATACGAATGATGCCAG 389
|||||

Qy 121 CTCGCTTGCGCCGGATATAAACCAACAGGACATCCAGATCAAGATGGAATGTC 180
|||||
Db 390 CTCGCTTGCGCCGGATATAAACCAACAGGACATCCAGATCAAGATGGAATGTC 449
|||||

RESULT 4
AM960685 553 bp mRNA linear EST 01-JUN-2000
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
EST.
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 553)
Begde, P., Qi, R., Abernathy, K., Dharap, S., Gaspar, R., Gay, C., Holt,
I. E., Saeed, A. I., Sharov, V., Lee, N. H., Yeatman, T. J. and
Quackenbush, J.
Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
Unpublished
Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: jquackenbush@igir.org
Plate: 149
Seq primer: Reverse.

FEATURES
source
Location/Qualifiers
1. .553
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_id="MAGE resequences, MAGF"
/note="Vector: pBluescriptSKm"

BASE COUNT
156 a 153 c 137 g 107 t

ORIGIN

Query Match 100.0%; Score 180; DB 10; Length 553;
Best Local Similarity 100.0%; Pred. No. 6.1e-45;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGAAGCTCCCTTTCTCAAGAAATGCCATCTGTGAACAGATGTAAGTCTCCAACTGT 60
|||||
Db 123 GGAAGCTCCCTTTCTCAAGAAATGCCATCTGTGAACAGATGTAAGTCTCCAACTGT 182
|||||

Qy 61 TCCAGATGTCACACCTGCTGCGGCACTGATGGGCTCACATATACGAATGATGCCAG 120
|||||
Db 183 TCCAGATGTCACACCTGCTGCGGCACTGATGGGCTCACATATACGAATGATGCCAG 242
|||||

Qy 121 CTCGCTTGCGCCGGATATAAACCAACAGGACATCCAGATCAAGATGGAATGTC 180
|||||
Db 243 CTCGCTTGCGCCGGATATAAACCAACAGGACATCCAGATCAAGATGGAATGTC 302
|||||

RESULT 5
BE543478 1115 bp mRNA linear EST 09-AUG-2000
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
EST.
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1115)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs.fda.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILN at:
http://image.llnl.gov
Plate: LHM8445 row: h column: 24
High quality sequence stop: 385.

FEATURES
source
Location/Qualifiers
1. .1115
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_id="IMAGE:3456839"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"
/clone_id="NIH-MGC_12"
/note="Organ: cervix; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.4 kb. Library prepared by Life
Technologies."

BASE COUNT
480 a 300 c 247 g 88 t

ORIGIN

Query Match 100.0%; Score 180; DB 10; Length 1115;
Best Local Similarity 100.0%; Pred. No. 7.8e-45;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGAAGCTCCCTTTCTCAAGAAATGCCATCTGTGAACAGATGTAAGTCTCCAACTGT 60
|||||
Db 72 GGAAGCTCCCTTTCTCAAGAAATGCCATCTGTGAACAGATGTAAGTCTCCAACTGT 131
|||||

Qy 61 TCCAGATGTCACACCTGCTGCGGCACTGATGGGCTCACATATACGAATGATGCCAG 120
|||||
Db 132 TCCAGATGTCACACCTGCTGCGGCACTGATGGGCTCACATATACGAATGATGCCAG 191
|||||

Qy 121 CTCGCTTGCGCCGGATATAAACCAACAGGACATCCAGATCAAGATGGAATGTC 180
|||||
Db 192 CTCGCTTGCGCCGGATATAAACCAACAGGACATCCAGATCAAGATGGAATGTC 251
|||||

RESULT 6
AA534438/c 321 bp mRNA linear EST 21-AUG-1997
LOCUS n18004.s1 NCI_CGAP_C03 Homo sapiens cDNA clone IMAGE:926191 3'
DEFINITION similar to SW:PE60_PIG P37109 PEPTIDE PEC-60 PRECURSOR. ;, mRNA
sequence.
ACCESSION AA534438
VERSION AA534438.1 GI:2278691
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 321)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbrr/image/image.html
Insert Length: 417 Std Error: 0.00
Seq primer: -40m13 fwd. RT from Amersham.
FEATURES
source
location/Qualifiers
1..321
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:926191"
/sex="pooled"
/tissue_type="colon"
/lab_host="DH10B"
/note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from 12 pooled bulk tumor samples and primed
with a Not I - oligo(dt) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified p773 vector. Library went through one round of
normalization."
BASE COUNT 62 a 83 c 81 g 95 t
ORIGIN
Query Match 93.0%; Score 167.4; DB 9; Length 321;
Best Local Similarity 98.9%; Pred. No. 4e-41;
Matches 179; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
1 GGAAGCTCCCTTCTCAAGATGCCATCTGTGAACACATGGTAGCTCCAACTGT 60
|||||
279 GGAAGCTCCCTTCTCAAGATGCCATCTGTGAACACATGGTAGCTCCAACTGT 220
|||||
61 TCCCAATGTCCAACTGTGTGGCGGCTGATGGGCTCACAATATGCAATG 119
|||||
219 TCCCAATGTCCAACTGTGTGGCGGCTGATGGGCTCACAATATGCAATG 160
|||||
120 GCTCTGCTGGCCGGGATAAAAACAAACAGACATCCAGATCATGGAATGGCAAT 179
|||||
159 GCTCTGCTGGCCGGGATAAAAACAAACAGACATCCAGATCATGGAATGGCAAT 100
|||||
180 C 180
1
99 C 99
RESULT 7

AT1732343/c 220 bp mRNA linear EST 13-DEC-1999
LOCUS n18004.x5 NCI_CGAP_C03 Homo sapiens cDNA clone IMAGE:926191 3'
DEFINITION similar to TR:060575 060575 GASTROINTESTINAL PEPTIDE. ;, mRNA
sequence.
ACCESSION AT1732343
VERSION AT1732343.1 GI:5053456
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 220)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbrr/image/image.html
FEATURES
source
location/Qualifiers
1..220
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:926191"
/sex="pooled"
/tissue_type="colon"
/lab_host="DH10B"
/note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from 12 pooled bulk tumor samples and primed
with a Not I - oligo(dt) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified p773 vector. Library went through one round of
normalization."
BASE COUNT 49 a 55 c 57 g 59 t
ORIGIN
Query Match 91.8%; Score 165.2; DB 9; Length 220;
Best Local Similarity 95.5%; Pred. No. 1.7e-40;
Matches 170; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
1 GGAAGCTCCCTTCTCAAGATGCCATCTGTGAACACATGGTAGCTCCAACTGT 60
|||||
178 GGAAGCTCCCTTCTCAAGATGCCATCTGTGAACACATGGTAGCTCCAACTGT 119
|||||
61 TCCCAATGTCCAACTGTGTGGCGGCTGATGGGCTCACAATATGCAATG 120
|||||
118 TCCCAATGTCCAACTGTGTGGCGGCTGATGGGCTCACAATATGCAATG 59
|||||
121 CTCTGCTGGCCGGGATAAAAACAAACAGACATCCAGATCATGGAATGGCAAT 178
|||||
58 CTCTGCTGGCCGGGATAAAAACAAACAGACATCCAGATCATGGAATGGCAAT 1

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 23, 2003, 10:26:50 ; Search time 783.841 Seconds

(without alignments)
9394.424 Million cell updates/sec

Title: US-09-941-309-1

Perfect score: 180
Sequence: 1 ggaagctcccttcctcaag.....tcataaagatgcaaatgc 180

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_da:*

2: gb_htg:*

3: gb_in:*

4: gb_ov:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vl:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htg_mus:*

34: em_htg_pln:*

35: em_htg_rdd:*

36: em_htg_mam:*

37: em_htg_vrt:*

38: em_sy:*

39: em_htgo_hum:*

40: em_htgo_mus:*

41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	180	100.0	180	6	AR166137 Sequence
2	180	100.0	180	6	E13090 Human an CD
3	180	100.0	386	9	AF048700 Homo sapi
4	180	100.0	388	6	AR028526 Sequence
5	180	100.0	388	6	AR075817 Sequence
6	180	100.0	388	6	BD007521 Tumor-ass
7	180	100.0	398	6	AR166138 Sequence
8	180	100.0	398	6	E13091 Human cDNA
9	180	100.0	402	6	AY014851 Sequence
10	127.2	70.7	364	4	S46866 PEC-60-gast
11	127.2	70.7	364	4	X67109 S.scrofa PE
12	120.8	67.1	515	10	Y11505 M.musculus
13	117.4	65.2	163396	9	AL161445 Human DNA
14	117.4	65.2	176584	2	AC027812 Homo sapi
15	117.4	65.2	183099	2	AC034144 Homo sapi
16	117.4	65.2	188540	2	AC011108 Homo sapi
17	81.2	45.1	192093	2	AC121205 Rattus no
18	81.2	45.1	230493	10	AL837521 Mouse DNA
19	81.2	45.1	242175	2	AC096463 Rattus no
20	81.2	23.0	168	6	I08731 Sequence 19
21	41.4	23.0	182	6	AI1350 Artificial
22	41.4	23.0	182	6	AI1352 Artificial
23	41.4	23.0	182	6	A24905 PSPI master
24	41.4	23.0	182	12	SYNPSTIA
25	41.4	23.0	186	6	A24907 PSPI master
26	41.4	22.1	482	6	E03479 cDNA encodi
27	39.8	22.1	5635	12	ASVPSKAN8
28	38.2	21.2	368	6	AX332609 Sequence
29	38.2	21.2	368	6	AX333248 Sequence
30	38.2	21.2	368	6	AX411203 Sequence
31	38.2	21.2	368	9	BD160949 Preventiv
32	38.2	21.2	368	9	HSPSTI
33	38.2	21.2	369	10	MMPIP12
34	36.6	20.3	179	6	E01725 Synthetic D
35	36.6	20.3	179	6	E02456 DNA encodin
36	36.6	20.3	204	12	SYNPSTIB
37	36.6	20.3	234	6	E01819 DNA sequenc
38	36.6	20.3	234	6	E03120 DNA sequenc
39	36.6	20.3	302	6	AR059723 Sequence
40	36.6	20.3	302	6	E08411 DNA sequenc
41	36.6	20.3	341	6	E01574 cDNA sequen
42	36.6	20.3	341	12	SYNECOPST
43	36.6	20.3	362	9	BC025790 Homo sapi
44	36.6	20.3	362	9	AR270687 Sequence
45	36.6	20.3	432	9	HUMPSSTI

ALIGNMENTS

RESULT 1

AR166137

LOCUS

DEFINITION Sequence 1 from patent US 6280968.

ACCESSION AR166137

VERSION AR166137.1

KEYWORDS GI:16241312

SOURCE

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 180)

AUTHORS Kato,S., Yamaguchi,T., Sekine,S. and Kamata,K.

TITLE Human PEC-60-like protein and DNA encoding the same

JOURNAL Patent: US 6280968-A 1 28-AUG-2001;

FEATURES Location/Qualifiers

PAT 17-OCT-2001

```
source 1. 180
BASE COUNT 53 a 49 c 40 g 38 t
ORIGIN
Query Match 100.0%; Score 180; DB 6; Length 180;
Best Local Similarity 100.0%; Pred. No. 2.6e-47;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAAGCTCCCTTTCACAGAAATGCCATCTGTGAACACATGTTAGAGTCTCCAACTGT 60
Db 1 GGAAGCTCCCTTTCACAGAAATGCCATCTGTGAACACATGTTAGAGTCTCCAACTGT 60
QY 61 TCCCAATGTCCAACTGCTGTGGCAGTGTGATGGCTCACAATACGAATGAATGCCAG 120
Db 61 TCCCAATGTCCAACTGCTGTGGCAGTGTGATGGCTCACAATACGAATGAATGCCAG 120
QY 121 CTCTGCTTGCGCCGGATAAAAACCAAGACATCCAGATCATGTAAGATGGCAATGC 180
Db 121 CTCTGCTTGCGCCGGATAAAAACCAAGACATCCAGATCATGTAAGATGGCAATGC 180

RESULT 2
LOCUS E13090 180 bp DNA linear PAT 27-APR-1998
DEFINITION Human an cDNA encoding a PEC-60-like protein.
ACCESSION E13090.1 GI:3251902
VERSION JP 1997124698-A/1.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1. (bases 1 to 180)
Kato, M., Yamaguchi, T., Sekine, S., and Kamata, M.
HUMAN PEC-60-LIKE PROTEIN AND DNA CAPABLE OF CODING THE SAME
PATENT: JP 1997124698-A 1 13-MAY-1997;
SAGAMI CHEM RES CENTER
OS Homo sapiens (human)
PN JP 1997124698-A/1
PD 13-MAY-1997
PE 27-OCT-1995 JP 1995280272
PI KATO MASASHI, YAMAGUCHI TOMOKO, SEKINE SHINGO, PI KAMATA
MITSUHIKA
PC C07K14/575, C07H21/04, C12N15/09//A61K38/22, A61K38/22,
PC C12P21/02, C12R1:19;
CC strandedness: Double;
CC topology: Linear;
FH Key Location/Qualifiers
FT source 1. 180
FT /organism="Homo sapiens",
FT /tissue_type="stomach cancer" FT mat_peptide
FT 1. 180
FT /product="PEC-60-like protein".
FEATURES
source 1. 180
Location/Qualifiers
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 53 a 49 c 40 g 38 t
ORIGIN
Query Match 100.0%; Score 180; DB 6; Length 180;
Best Local Similarity 100.0%; Pred. No. 2.6e-47;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAAGCTCCCTTTCACAGAAATGCCATCTGTGAACACATGTTAGAGTCTCCAACTGT 60
Db 1 GGAAGCTCCCTTTCACAGAAATGCCATCTGTGAACACATGTTAGAGTCTCCAACTGT 60
QY 61 TCCCAATGTCCAACTGCTGTGGCAGTGTGATGGCTCACAATACGAATGAATGCCAG 120
Db 61 TCCCAATGTCCAACTGCTGTGGCAGTGTGATGGCTCACAATACGAATGAATGCCAG 120
QY 121 CTCTGCTTGCGCCGGATAAAAACCAAGACATCCAGATCATGTAAGATGGCAATGC 180
Db 121 CTCTGCTTGCGCCGGATAAAAACCAAGACATCCAGATCATGTAAGATGGCAATGC 180

RESULT 4
LOCUS AR028526 388 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 2 from patent US 5858710.
ACCESSION AR028526
VERSION AR028526.1 GI:5940499
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
```

```
Db 61 TCCCAATGTCCAACTGCTGTGGCAGTGTGATGGCTCACAATACGAATGAATGCCAG 120
QY 121 CTCTGCTTGCGCCGGATAAAAACCAAGACATCCAGATCATGTAAGATGGCAATGC 180
Db 121 CTCTGCTTGCGCCGGATAAAAACCAAGACATCCAGATCATGTAAGATGGCAATGC 180

RESULT 3
LOCUS AF048700 386 bp mRNA linear PRI 06-MAR-1998
DEFINITION Homo sapiens gastrointestinal peptide (PEC-60) mRNA, complete cds.
ACCESSION AF048700.1 GI:2935439
VERSION AF048700.1
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1. (bases 1 to 386)
Hu, G.
Human gastrointestinal peptide mRNA, homolog of swine PEC-60
JOURNAL Unpublished
AUTHORS Hu, G.
TITLE Direct Submission
REFERENCE 2 (bases 1 to 386)
Hu, G.
Submitted (17-FEB-1998) Shanghai Institute of Cell Biology, 320
Yue-Yang Road, Shanghai 200031, China
LOCATION/Qualifiers
source 1. 386
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1. 386
/gene="PEC-60"
12. 272
/gene="PEC-60"
/note="similar to Sus scrofa PEC-60"
/codon_start=1
/product="gastrointestinal peptide"
/protein_id="AAC05124.1"
/db_xref="GI:2935440"
/translacion="MAVRQWVIALALALLVDREVPVAGKLPFSRMPICHEWESP
TCSQMNLCVGTGLTYTNECOLCLARITKDIQIMKDKC"
BASE COUNT 119 a 97 c 98 g 72 t
ORIGIN
Query Match 100.0%; Score 180; DB 9; Length 386;
Best Local Similarity 100.0%; Pred. No. 2.6e-47;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAAGCTCCCTTTCACAGAAATGCCATCTGTGAACACATGTTAGAGTCTCCAACTGT 60
Db 90 GGAAGCTCCCTTTCACAGAAATGCCATCTGTGAACACATGTTAGAGTCTCCAACTGT 149
QY 61 TCCCAATGTCCAACTGCTGTGGCAGTGTGATGGCTCACAATACGAATGAATGCCAG 120
Db 150 TCCCAATGTCCAACTGCTGTGGCAGTGTGATGGCTCACAATACGAATGAATGCCAG 209
QY 121 CTCTGCTTGCGCCGGATAAAAACCAAGACATCCAGATCATGTAAGATGGCAATGC 180
Db 210 CTCTGCTTGCGCCGGATAAAAACCAAGACATCCAGATCATGTAAGATGGCAATGC 269
```

Unclassified.
REFERENCE 1 (bases 1 to 388)
AUTHORS Bandman,O., Goli,S.K. and Murry,L.E.
TITLE Tumor-associated kazal inhibitor-like polypeptides and encoding polynucleotides
JOURNAL Patent: US 5858210-A 2 12-JAN-1999;
FEATURES Location/Qualifiers
source 1..388
/organism="unknown"
BASE COUNT 105 a 105 c 102 g 76 t
ORIGIN
Query Match 100.0%; Score 180; DB 6; Length 388;
Best Local Similarity 100.0%; Pred. No. 2.6e-47;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5
AR075817
LOCUS AR075817 388 bp DNA linear PAT 30-AUG-2000
DEFINITION Sequence 2 from patent US 5958699.
ACCESSION AR075817
VERSION AR075817.1 GI:10002563
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 388)
AUTHORS Bandman,O., Goli,S.K. and Murry,L.E.
TITLE Tumor-associated Kazal Inhibitor
JOURNAL Patent: US 5958699-A 2 28-SEP-1999;
FEATURES Location/Qualifiers
source 1..388
/organism="unknown"
BASE COUNT 105 a 105 c 102 g 76 t
ORIGIN
Query Match 100.0%; Score 180; DB 6; Length 388;
Best Local Similarity 100.0%; Pred. No. 2.6e-47;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 6
BD007521
LOCUS BD007521 388 bp DNA linear PAT 31-JAN-2002
DEFINITION Tumor-associated KAZAL type inhibitor.
ACCESSION BD007521
VERSION BD007521.1 GI:18635894
KEYWORDS JP 2001503629-A/1.

SOURCE unidentified
ORGANISM unidentified
REFERENCE 1 (bases 1 to 388)
AUTHORS Bandman,O., Goli,S.K. and Murry,L.E.
TITLE Tumor-associated KAZAL type inhibitor
JOURNAL Patent: JP 2001503629-A 1 21-MAR-2001;
COMMENT INCYTE PHARMACEUTICALS INC
OS Unidentified
PN JP 2001503629-A/1
PD 21-MAR-2001
PF 05-NOV-1997 JP 1998521770
PI 06-NOV-1996 US 08/744670
PC OLGA BANDMAN,SURYA K GOLI,LYNN E MURRY
PC C12N15/09,A61K38/55,A61K39/395,A61P1/00,A61P35/00,C07K14/81,
PC C07K16/38,
PC C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12Q1/68,G01N33/574, PC
C12N15/00,C12N5/00,
PC A61K37/64
CC
FH Key
FT source
FT Location/Qualifiers
FEATURES 1..388
source /organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
BASE COUNT 105 a 105 c 102 g 76 t
ORIGIN
Query Match 100.0%; Score 180; DB 6; Length 388;
Best Local Similarity 100.0%; Pred. No. 2.6e-47;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 7
AR166138
LOCUS AR166138 398 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 2 from patent US 6280968.
ACCESSION AR166138
VERSION AR166138.1 GI:16241313
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 398)
AUTHORS Kato,S., Yamaguchi,T., Sekine,S. and Kamata,K.
TITLE Human PEC-60-like protein and DNA encoding the same
JOURNAL Patent: US 6280968-A 2 28-AUG-2001;
FEATURES Location/Qualifiers
source 1..398
/organism="unknown"
BASE COUNT 107 a 110 c 105 g 76 t
ORIGIN
Query Match 100.0%; Score 180; DB 6; Length 398;
Best Local Similarity 100.0%; Pred. No. 2.6e-47;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 8
GGAAAGCTCCCTTTCTCAAGAAATGCCCATCTGTGAACACATGTCAGTCTCCAACTGT 60
1 GGAAGCTCCCTTTCTCAAGAAATGCCCATCTGTGAACACATGTCAGTCTCCAACTGT 60
111 GGAAGCTCCCTTTCTCAAGAAATGCCCATCTGTGAACACATGTCAGTCTCCAACTGT 170
61 TCCAGATGTCACACCTGCTGCGGCACTGATGGGCTCACATATACGAATGATGCCAG 120
171 TCCAGATGTCACACCTGCTGCGGCACTGATGGGCTCACATATACGAATGATGCCAG 230
121 CTCTGCTTTGGCCCGGATAAAAACCAACAGACATCCAGATCATGAAAGATGGCAATGC 180
231 CTCTGCTTTGGCCCGGATAAAAACCAACAGACATCCAGATCATGAAAGATGGCAATGC 290

Db	121	GGAAAGCTCCCTTCTCTCAAGATGCCATCTGTGACACATGTRGATCTCCAACTGT	180
QY	61	TCCAGATGTCACACCTGTGTCGGGCACTGATGGCTCACAATATAGCAATGATGCCAG	120
Db	181	TCCAGATGTCACACCTGTGTCGGGCACTGATGGCTCACAATATAGCAATGATGCCAG	240
QY	121	CTGCGCTGGCCCGGATATAAAACCAACAGACATCCAGATCATGAAGATGGCAAAATGC	180
Db	241	CTGCGCTGGCCCGGATATAAAACCAACAGACATCCAGATCATGAAGATGGCAAAATGC	300
RESULT 8			
LOCUS	E13091	398 bp	DNA linear PAT 27-APR-1998
DEFINITION	Human cDNA encoding a PEC-60-like protein.		
ACCESSION	E13091		
VERSION	E13091.1	GI:3251903	
KEYWORDS	JP 1997124698-A/2.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Kato,M., Yamaguchi,T., Sekine,S. and Kamata,M.		
TITLE	HUMAN PEC-60-LIKE PROTEIN AND DNA CAPABLE OF CODING THE SAME		
JOURNAL	Patent: JP 1997124698-A 2 13-MAY-1997;		
COMMENT	SAGAMI CHEM RES CENTER		
	OS Homo sapiens (human)		
	PN JP 1997124698-A/2		
	PD 13-MAY-1997		
	PI 27-OCT-1995 JP 1995280272		
	PI KATO MASASHI, YAMAGUCHI TOMOKO, SEKINE SHINGO, PI KAMATA MITSURUHA		
	PC C07K14/575, C07H21/04, C12M15/09//A61K38/22, A61K38/22, A61K38/22		
	PC C12P21/02,		
	PC (C12P21/02, C12R1:19);		
	CC strandedness: Double;		
	CC topology: Linear;		
	PH key	Location/Qualifiers	
	FT source	1..398	
	FT	/organism='Homo sapiens'	
	FT	/tissue_type='stomach cancer' FT CDS	
	FT	43..303	
	FT	sig_peptide 43..120	
	FT	mat_peptide 121..300	
	FT	/product='PEC-60-like protein',	
FEATURES	Location/Qualifiers		
source	1..398		
	/organism='Homo sapiens'		
	/mol_type='genomic DNA'		
	/db_xref='taxon:9606'		
BASE COUNT	107 a 110 c 105 g 76 t		
ORIGIN			
Query Match	100.0%; Score 180; DB 6; Length 398;		
Best Local Similarity	100.0%; Pred. No. 2.6e-47;		
Matches	180; Conservative 100.0%; Mismatches 0; Indels 0; Gaps 0.0;		
QY	1	GGAAAGCTCCCTTCTCTCAAGATGCCATCTGTGACACATGTRGATCTCCAACTGT	60
Db	121	GGAAAGCTCCCTTCTCTCAAGATGCCATCTGTGACACATGTRGATCTCCAACTGT	180
QY	61	TCCAGATGTCACACCTGTGTCGGGCACTGATGGCTCACAATATAGCAATGATGCCAG	120
Db	181	TCCAGATGTCACACCTGTGTCGGGCACTGATGGCTCACAATATAGCAATGATGCCAG	240
QY	121	CTGCGCTGGCCCGGATATAAAACCAACAGACATCCAGATCATGAAGATGGCAAAATGC	180
Db	241	CTGCGCTGGCCCGGATATAAAACCAACAGACATCCAGATCATGAAGATGGCAAAATGC	300

[illegible]

Db 302 CTTGCTGACCGGATGAAACCATGAGACATTCATCAATCATGAGAGTGGCCATGCG 361
 RESULT 13
 AL161445
 LOCUS
 DEFINITION AL161445 163396 bp DNA linear PRI 04-DEC-2001
 Human DNA sequence from clone RP11-326F20 on chromosome 9. Contains
 the GGT1 gene for glycoprotein alpha-galactosyl transferase 1, the
 SPINK4 gene for Kazal type 4 serine protease inhibitor, the 3' end
 of the BAG1 gene for BCL2-associated athanogene and a Cpg island,
 complete sequence.
 AL161445
 VERSION AL161445.10 GI:11137678
 KEYWORDS HFG; athanogene; BAG1; BCL2; Cpg island; galactosyltransferase;
 GGT1; glycoprotein; Kazal; serine protease inhibitor; SPINK4.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 163396)
 AUTHORS Ramsay, H.
 JOURNAL Direct Submission
 Submitted (04-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
 On Nov 10, 2000 this sequence version replaced gi:11071610.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
 Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr9
 This sequence is the entire insert of clone RP11-326F20. The true
 left end of clone RP11-344B24 is at 100366 in this sequence. The
 true right end of clone RP11-54K16 is at 54767 in this sequence.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. RP11-326F20 is from
 the library RPCT-11.2 constructed by the group of Pieter de Jong.
 For further details see
 http://www.chori.org/bacpac/home.htm
 VECTOR: pBACe3.6.
 FEATURES
 source
 Location/Qualifiers
 1. 163396
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="9"
 /clone="RP11-326F20"
 /clone_lib="RPCT-11.2"
 5. 237
 repeat_region
 /note="AluY repeat: matches 77. .309 of consensus"
 239. 314
 repeat_region
 /note="L1M4 repeat: matches 2591. .2666 of consensus"
 331. 438
 repeat_region
 /note="MER44A repeat: matches 225. .331 of consensus"
 440. 555
 repeat_region
 /note="AluSq/x repeat: matches 11. .126 of consensus"
 566. 708
 repeat_region

/note="AluJo/FRAM repeat: matches 158. .298 of consensus"
 795. 1193
 repeat_region
 /note="L1M4 repeat: matches 3657. .4070 of consensus"
 1194. 1491
 repeat_region
 /note="AluSp repeat: matches 1. .296 of consensus"
 1492. 1680
 repeat_region
 /note="L1M4 repeat: matches 4070. .4265 of consensus"
 1681. 1981
 repeat_region
 /note="AluSx repeat: matches 1. .301 of consensus"
 1982. 2181
 repeat_region
 /note="L1M4 repeat: matches 4265. .4471 of consensus"
 2182. 2486
 repeat_region
 /note="Aluub repeat: matches 1. .304 of consensus"
 2487. 2783
 repeat_region
 /note="L1M4 repeat: matches 4471. .4701 of consensus"
 3046. 3192
 repeat_region
 /note="AluSq/x repeat: matches 1. .147 of consensus"
 3196. 3247
 repeat_region
 /note="MIR repeat: matches 17. .66 of consensus"
 4058. 4566
 repeat_region
 /note="MER41B repeat: matches 150. .635 of consensus"
 4567. 4862
 repeat_region
 /note="AluSq repeat: matches 1. .300 of consensus"
 4863. 5008
 repeat_region
 /note="MER41B repeat: matches 1. .150 of consensus"
 6433. 6490
 repeat_region
 /note="MIR repeat: matches 81. .138 of consensus"
 7523. 7596
 repeat_region
 /note="L1P1 repeat: matches 5753. .5825 of consensus"
 8418. 8461
 repeat_region
 /note="L1MB4 repeat: matches 6132. .6184 of consensus"
 8462. 8769
 repeat_region
 /note="AluSx repeat: matches 1. .304 of consensus"
 8770. 8783
 repeat_region
 /note="L1MB4 repeat: matches 6145. .6132 of consensus"
 8784. 9085
 repeat_region
 /note="AluSp repeat: matches 1. .302 of consensus"
 9086. 9106
 repeat_region
 /note="L1MB4 repeat: matches 6122. .6146 of consensus"
 9152. 9751
 repeat_region
 /note="L1MB3 repeat: matches 5512. .6125 of consensus"
 10464. 10534
 repeat_region
 /note="MIR repeat: matches 61. .130 of consensus"
 11734. 11790
 repeat_region
 /note="L12 repeat: matches 2688. .2742 of consensus"
 11747. 11869
 repeat_region
 /note="MIR repeat: matches 132. .252 of consensus"
 11874. 12110
 repeat_region
 /note="FLAMC repeat: matches 1. .134 of consensus"
 12121. 12219
 repeat_region
 /note="MIR repeat: matches 35. .142 of consensus"
 12243. 12346
 repeat_region
 /note="52 copies 2 mer tg 71% conserved"
 12254. 12343
 repeat_region
 /note="9 copies 10 mer ggtgtgtgt 72% conserved"
 complement(12461. .12790)
 /note="match: GSS: Em:AO210501"
 complement(12571. .12678)
 /note="match: GSS: Em:AO177335"
 12602. 13170
 repeat_region
 /note="MER21B repeat: matches 232. .794 of consensus"
 13171. 13289
 repeat_region
 /note="FLAM A repeat: matches 1. .122 of consensus"
 13290. 13517
 repeat_region
 /note="MER21B repeat: matches 4. .232 of consensus"
 14033. 14159
 repeat_region
 /note="MIR repeat: matches 112. .257 of consensus"
 complement(14283. .70891)
 /gene="GGTA1"
 mRNA
 complement(join(14283. 17232,17420. 17524,19637. 19759,
 24065. 24252,38835. 39070,70404. 70891))
 /gene="GGTA1"
 /product="ba326F20.1 (glycoprotein, alpha-galactosyl

```
transferase 1)"
/note="glycoprotein, alpha-galactosyl transferase 1
match: cDNAs: Em:J03880 Em:D00314 Em:E02227 Em:M13569
Em:X14085 Em:A23697 Em:X14558 Em:M22921 Em:U10473
Em:E02228 Em:X13223 Em:D29605 Em:M13214 Em:X55415
Em:U19890 Em:Y12510 Em:M13701 Em:U19889 Em:AF038660
Em:U10474 Em:AB024434 Em:AB019541 Em:U10472 Em:AF142670
match: ESTs: Em:BF124086 Em:AK467118 Em:AA769459
Em:AA476930 Em:AI192425 Em:AI572046 Em:AI193377
Em:AA830559 Em:AA889902 Em:AI028425 Em:AW362033
Em:AI149203 Em:AI565973"
/evidence=not_experimental
complement(14283)
/gene="GCTAI"
complement(14305..14310)
/gene="GCTAI"
14897..14934
/note="19 copies 2 mer aa 92% conserved"
complement(15755..17099)
/gene="GCTAI"
/note="match: STS: Em:C05861"
complement(16558..16912)
/gene="GCTAI"
/note="match: STS: Em:G59840"
complement(join(17100..17232,17420..17524,19637..19759,
24065..24252,38835..39070,70404..70815))
/gene="GCTAI"
/note="match: proteins: Tr:O60512"
/codon_start=1
/evidence=not_experimental
/product="ba326f20.1 (glycoprotein, alpha-galactosyl
transferase 1)"
/protein_id="CAD13306.1"
/ID_xref="GI:17384434"
/translation="MRLEPLLSGSAMPASLQACRLNVAICALHLYLVLYYLNG
RDLRLPOLVGVSTPLGGSSAAIGSGSELTGGRPPGLGASOPRGDSSP
VVDGSGPASNLTSPVPHHTALSIPACPESEPLVGPMLIEFNPMDLVKKNPN
VKMGGRVAPRDQVSPKXVATITPERNOEHLKYVIYHPLVROODVCIYINAG
DITFNRKALLNVEFOELAKDYDTITCFYSVDILIPMDHNAIKCFSPRISVAMDF
GFSLPYVQIFGVGSALSKDQFLITNGPNMYWGGEDDIFNRLVFRGMSISRPNV
VGRCRMIRHSRDKKNSEPNQRFDRIRIAHTKETMLSDGLNLTGYLDVQRYPLVYQITIV
DIGTPS"
repeat_region
17946..18109
/note="12 repeat: matches 2241..2411 of consensus"
19126..19243
/note="LIMC5 repeat: matches 7805..7927 of consensus"
repeat_region
19877..20157
/note="AluSc repeat: matches 11..299 of consensus"
20168..20468
/note="AluSx repeat: matches 1..301 of consensus"
22027..22341
/note="AluJo repeat: matches 4..309 of consensus"
22507..22807
/note="AluSx repeat: matches 1..301 of consensus"
23081..23376
/note="AluJo repeat: matches 1..297 of consensus"
23549..23852
/note="AluJo repeat: matches 1..305 of consensus"
24409..24625
/note="L1ME repeat: matches 5615..5826 of consensus"

Query Match 65.2% Score 117.4; DB: 9; Length 163396;
Best Local Similarity 91.9% P-Val: 5.3e-27;
Matches 124; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
```

```
QY 7 CTCCTCTTCTCAGAAATGCCATCTGTGACACATGTAGAGTCCACCTGTTCACG 66
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 150244 CTCCTCTCTTCCACAGCCCATCTGTGACACATGTAGAGTCTCAACTGTTCACG 150303

QY 67 ATGTCAACCTGTGTGCGGCGACTGATGGCTCACAATATAGATGAATGCCACTTGC 126
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 150304 ATGTCAACCTGTGTGCGGCGACTGATGGCTCACAATATAGATGAATGCCACTTGC 150363
```

QY 127 TTGGCCCGATATAA 141
Db 150364 TTGGCCCGATATAA 150378

RESULT 14
AC027812/c
LOCUS
DEFINITION
Homo sapiens chromosome 9 clone RP11-366f19 map 9, WORKING DRAFT
SEQUENCE, 29 unordered pieces.
AC027812
VERSION
AC027812.2 GI:7677927
HTG: HTGS PHASE1: HTGS_DRAFT.
SOURCE
Homo sapiens (human)
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavskiy, L., Bouckhgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choquel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J., S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J., C., Iliev, I., Johnson, R., Jones, C., Kann, L., Kartas, A., Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Liu, G., Locke, K., MacDonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheters, R., Meldrum, J., Meneses, L., Minova, T., Miranda, C., Mlenka, V., Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Plerre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rotman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tefaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE
JOURNAL
REFERENCE
AUTHORS
Direct Submission
Submitted (01-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 176584)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavskiy, L., Bouckhgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choquel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J., S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J., C., Iliev, I., Johnson, R., Jones, C., Kann, L., Kartas, A., Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Liu, G., Locke, K., MacDonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheters, R., Meldrum, J., Meneses, L., Minova, T., Miranda, C., Mlenka, V., Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Plerre, N., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rotman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tefaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT
Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 2, 2000 this sequence version replaced g1:7382654.
All repeats were identified using RepeatMasker:

OY 127 TTGGCCCGGATATAA 141
DB 136140 TTGGCCCGGATATAA 136126

RESULT 15
AC034144
LOCUS
DEFINITION Homo sapiens chromosome 9 clone RP11-126M6, WORKING DRAFT SEQUENCE,
5 unordered pieces.
AC034144
AC034144
VERSION AC034144.5 GI:9929777
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 183099)
Waterston, R.H.
The sequence of Homo sapiens clone
2 (bases 1 to 183099)
Waterston, R.H.
Direct Submission
Submitted (04-APR-2000) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Aug 26, 2000 this sequence version replaced gi:9795673.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0126M06
----- Summary Statistics -----
Sequencing vector: plasmid; 08
Chemistry: Dye-Primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 179431 bases at least Q40
Consensus quality: 180878 bases at least Q30
Consensus quality: 181543 bases at least Q20
Insert size: 183000; agarose-fp
Insert size: 182699; sum-of-contigs
Quality coverage: 6.27 in Q20 bases; agarose-fp
Quality coverage: 6.20 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 24313: contig of 24313 bp in length
* 24314 24413: gap of unknown length
* 24414 54520: contig of 30107 bp in length
* 54521 54620: gap of unknown length
* 54621 93200: contig of 38580 bp in length
* 93201 93300: gap of unknown length
* 93301 132512: contig of 39212 bp in length
* 132513 132613: gap of unknown length
* 132613 183099: contig of 50487 bp in length.
Location/Qualifiers
1. 183099
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="9"

misc_feature
1..24313
/clone="RP11-126M6"
misc_feature
24414..54520
/note="assembly_name:Contig16"
misc_feature
54621..93200
/note="assembly_name:Contig17"
misc_feature
93301..132512
/note="assembly_name:Contig18"
misc_feature
132613..183099
/note="assembly_name:Contig19"
BASE COUNT 52286 a 40806 c 40143 g 49464 t 400 others
ORIGIN

Query Match 65.2%; Score 117.4; DB 2; Length 183099;
Best Local Similarity 91.9%; Pred. No. 5.3e-27;
Matches 124; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 7 CTCCCTTCACAGATGCCCATCTGTGACACATGTAGAGTCTCAACCTGTTCCAG 66
||||| |
DB 64694 CTCCCTTCCTCCACAGACCCATCTGTGAACATGTAGAGTCTCAACCTGTTCCAG 64753
||||| |

OY 67 ATGTCCAACTGCTGTGGGACGTGATGGGCTCACATATACGAATGAATGCCAGCTTGC 126
||||| |

DB 64754 ATGTCCAACTGCTGTGGGACGTGATGGGCTCACATATACGAATGAATGCCAGCTTGC 64813
||||| |

OY 127 TTGGCCCGGATATAA 141
||||| |

DB 64814 TTGGCCCGGATATAA 64828
||||| |

Search completed: September 23, 2003, 12:11:12
Job time : 787.841 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 23, 2003, 09:52:25 ; Search time 124.256 Seconds
(without alignments)
3910.464 Million cell updates/sec

Title: US-09-941-309-1

Perfect score: 180
Sequence: 1 ggaagctcccttctcaag.....tcattgaagatgccaatgc 180

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_19Jun03:*

- 1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
- 23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
- 25: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	180	100.0	323	24	ABV86895	Human colon cancer
2	180	100.0	323	24	ABV89019	Human colon cancer
3	180	100.0	389	19	AAV38073	Human tumour-associ
4	180	100.0	394	24	ABV97573	Human pancreatic c
5	180	100.0	398	18	AAV64673	Human PEC-60-type
6	180	100.0	402	20	AAZ77490	Human ovarian tumo
7	180	100.0	433	21	AAC98176	Human colon cancer
8	180	100.0	433	22	AAH34805	Human colon cancer

9	151.8	84.3	468	24	ABO60075	Human colon cancer
10	137.4	76.3	283	24	ABO60577	Human colon cancer
11	44.6	24.8	168	9	AAH80453	Sequence encoding
12	41.4	23.0	168	9	AAH80452	Sequence encoding
13	41.4	23.0	182	9	AAH80030	Synthetic DNA enco
14	41.4	23.0	182	9	AAH82216	Synthetic DNA enco
15	41.4	23.0	482	13	AAO24083	Secretion plasmid
16	39.8	22.1	182	9	AAH82217	Synthetic DNA enco
17	39.8	22.1	182	9	AAH82232	Synthetic DNA enco
18	39.8	22.1	415	12	AAQ12109	Human pancreatic s
19	39.8	22.1	415	17	AAH15982	PSKAN8 fragment #1
20	38.8	21.6	182	9	AAH82227	Synthetic DNA enco
21	38.8	21.6	182	9	AAH82228	Synthetic DNA enco
22	38.8	21.6	182	9	AAH82233	Synthetic DNA enco
23	38.2	21.2	171	11	AAQ01494	Modified human pan
24	38.2	21.2	182	9	AAH82221	Synthetic DNA enco
25	38.2	21.2	368	24	ABK6355	Human pancreas sec
26	38.2	21.2	368	24	ABN97352	Gene #3850 used to
27	38.2	21.2	368	24	ABL64781	Lung cancer relate
28	38.2	21.2	368	24	ABL65420	Lung cancer relate
29	37.4	20.8	182	9	AAH82224	Synthetic DNA enco
30	37.4	20.8	182	9	AAH82225	Synthetic DNA enco
31	37.4	20.8	182	9	AAH82226	Synthetic DNA enco
32	37.4	20.8	182	9	AAH82223	Synthetic DNA enco
33	37.2	20.7	182	9	AAH82229	Synthetic DNA enco
34	37.2	20.7	182	9	AAH82230	Synthetic DNA enco
35	36.6	20.3	171	11	AAQ01577	Modified human pan
36	36.6	20.3	179	9	AAH81535	Human pancreatic s
37	36.6	20.3	182	9	AAH82222	Synthetic DNA enco
38	36.6	20.3	200	10	AAH91164	Lambda phage cti s
39	36.6	20.3	200	12	AAQ14261	Pancreas secretory
40	36.6	20.3	234	10	AAH91165	Sequence encoding
41	36.6	20.3	234	12	AAQ14262	Pancreas secretory
42	36.6	20.3	238	18	AAH84523	hprt1.M2 gene for
43	36.6	20.3	302	15	AAQ78611	HindIII-BamHI frag
44	36.6	20.3	341	9	AAH80460	Human pancreatic s
45	36.6	20.3	362	24	ABO60575	Human colon cancer

ALIGNMENTS

RESULT 1
ID ABV86895 standard; cDNA: 323 BP.
AC ABV86895;
XX
XX 13-DEC-2002 (first entry)
DE Human colon cancer related cDNA SEQ ID NO 206.
XX
XX Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine; gene;
KW ss.
XX Homo sapiens.
OS
XX WO200258334-A2.
PN
XX 01-AUG-2002.
PD
XX 19-NOV-2001; 2001MO-US43704.
FE
XX 20-NOV-2000; 2000US-252222P.
PR
XX 06-FEB-2001; 2001US-267011P.
PR
XX 28-MAR-2001; 2001US-279670P.
PR
XX 10-JUL-2001; 2001US-304037P.
PA
XX (CORI-) CORIXA CORP.
PI
XX Stolk JA, Xu J, Chenaule RA, Meagher MJ, Secretist H, King GE;
XX WPI; 2002-608400/65.
DR

XX New isolated tumor colon polynucleotide and polypeptide, useful for the
 PT diagnosis, prevention and/or treatment of cancer, in particular colon
 PT cancer
 XX
 PS Claim 1: SEQ ID NO 206; 266pp + Sequence Listing: English.
 XX
 CC The invention relates to a human colon tumour expressed polynucleotide
 CC (I) encoding a polypeptide (II, ABP67991-ABP67996) comprising: (i) any of
 CC 2600 fully defined nucleotide sequences (ABV8669-ABV89289); (ii)
 CC complements of (i); (iii) at least 20 contiguous residues of (i); (iv)
 CC sequences that hybridize to (i), under moderately stringent conditions;
 CC (v) sequences having at least 75% or 90% identity to (i); or (vi)
 CC degenerate variants of (i). The compositions and methods of the present
 CC invention are useful for the diagnosis, prevention and/or treatment of
 CC cancer, particularly colon cancer. (i) can be used in gene therapy and
 CC (i) and (ii) are useful in pharmaceutical compositions such as vaccines.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.

Sequence 323 BP; 66 A; 81 C; 93 G; 83 T; 0 other;

Query Match 100.0%; Score 180; DB 24; Length 323;
 Best Local Similarity 100.0%; Pred. No. 1.1e-51;
 Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAAAGCTCCCTTCTCAAGAAATGCCATCTGTGAACACATGGTAGAGTCTCCAACTGT 60
 DB 217 GGAAAGCTCCCTTCTCAAGAAATGCCATCTGTGAACACATGGTAGAGTCTCCAACTGT 158
 QY 61 TCCCAATATCCAACTGTCTGGCGGCTGATGGGCTCACATATACGAATGAATCCAG 120
 DB 157 TCCCAATATCCAACTGTCTGGCGGCTGATGGGCTCACATATACGAATGAATCCAG 98
 QY 121 CTCTGCTTGGCCCGGATTAACCAACAGACATCAGATCATGAAGATGGCAATGC 180
 DB 97 CTCTGCTTGGCCCGGATTAACCAACAGACATCAGATCATGAAGATGGCAATGC 38

RESULT 2
 ABV89019
 ID ABV89019 standard; cDNA; 323 BP.
 XX
 AC ABV89019;
 XX
 DT 13-DEC-2002 (first entry)
 XX
 XX Human colon cancer related cDNA SEQ ID NO 2330.

Human colon cancer; cancer; cytostatic; tumour; gene therapy; vaccine; gene;
 Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine; gene;
 ss.
 Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine; gene;
 ss.
 Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine; gene;
 ss.

Human sapiens.
 WO200258534-A2.
 PD 01-AUG-2002.
 XX
 PD 19-NOV-2001; 2001WO-US43704.
 XX
 PF 20-NOV-2000; 2000US-252222P.
 XX
 PR 06-FEB-2001; 2001US-267011P.
 XX
 PR 28-MAR-2001; 2001US-279670P.
 XX
 PR 10-JUL-2001; 2001US-304037P.
 XX

(CORI-) CORIXA CORP.

Stolk JA, Xu J, Chenault RA, Meagher MJ, Secretist H, King GE;
 WPI, 2002-608400/65.

New isolated tumor colon polynucleotide and polypeptide, useful for the

PT diagnosis, prevention and/or treatment of cancer, in particular colon
 PT cancer
 XX
 PS Claim 1: SEQ ID NO 2330; 266pp + Sequence Listing: English.
 XX
 CC The invention relates to a human colon tumour expressed polynucleotide
 CC (I) encoding a polypeptide (II, ABP67991-ABP67996) comprising: (i) any of
 CC 2600 fully defined nucleotide sequences (ABV8669-ABV89289); (ii)
 CC complements of (i); (iii) at least 20 contiguous residues of (i); (iv)
 CC sequences that hybridize to (i), under moderately stringent conditions;
 CC (v) sequences having at least 75% or 90% identity to (i); or (vi)
 CC degenerate variants of (i). The compositions and methods of the present
 CC invention are useful for the diagnosis, prevention and/or treatment of
 CC cancer, particularly colon cancer. (i) can be used in gene therapy and
 CC (i) and (ii) are useful in pharmaceutical compositions such as vaccines.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.

Sequence 323 BP; 83 A; 93 C; 81 G; 66 T; 0 other;

Query Match 100.0%; Score 180; DB 24; Length 323;
 Best Local Similarity 100.0%; Pred. No. 1.1e-51;
 Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAAAGCTCCCTTCTCAAGAAATGCCATCTGTGAACACATGGTAGAGTCTCCAACTGT 60
 DB 107 GGAAAGCTCCCTTCTCAAGAAATGCCATCTGTGAACACATGGTAGAGTCTCCAACTGT 166
 QY 61 TCCCAATATCCAACTGTCTGGCGGCTGATGGGCTCACATATACGAATGAATCCAG 120
 DB 167 TCCCAATATCCAACTGTCTGGCGGCTGATGGGCTCACATATACGAATGAATCCAG 226
 QY 121 CTCTGCTTGGCCCGGATTAACCAACAGACATCAGATCATGAAGATGGCAATGC 180
 DB 227 CTCTGCTTGGCCCGGATTAACCAACAGACATCAGATCATGAAGATGGCAATGC 286

RESULT 3
 AAV38073
 ID AAV38073 standard; cDNA; 389 BP.
 XX
 AC AAV38073;
 XX
 DT 14-SEP-1998 (first entry)
 XX
 XX Human tumour-associated Kazal inhibitor encoding cDNA.

Human tumour-associated Kazal inhibitor; TAKI; inflammation; cancer;
 ulcerative colitis; Crohn's disease; inflammatory cytokine; elastase;
 serine protease; glycoprotein hormone; trypsin; chymotrypsin; pepsin; ss.

Human sapiens.
 WO9820132-A1.
 PD 14-MAY-1998.
 XX
 PF 05-NOV-1997; 97WO-US20204.
 XX
 PR 06-NOV-1996; 96US-0744670.
 XX

Key Location/Qualifiers
 FT CDS 34..294
 FT /tag=a
 FT /product="tumour-associated Kazal inhibitor"

(INCY-) INCYTE PHARM INC.

Bandman O, Goli SK, Murry LE;

WPI, 1998-286948/25.
 P-PSDB; AAW62074.

XX New isolated tumour-associated Kazal inhibitor, TAKI - used to
 PT develop products for the diagnosis and treatment of e.g.
 PT inflammation, ulcerative colitis, Crohn's disease or cancer
 XX
 PS Claim 5; Fig 1; 60pp; English.
 XX
 CC The present sequence encodes human tumour-associated Kazal inhibitor
 CC (TKI). TAKI is expressed in response to inflammatory cytokines and
 CC functions in the inhibition of serine proteases or glycoprotein hormones
 CC present in diseased tissues. The TAKI can be used to inhibit the tissue
 CC destruction associated with the production of excess proteases such as
 CC trypsin, chymotrypsin, elastase or pepsin. Products from the present
 CC invention can be used in the treatment of e.g. proliferative cell
 CC division in inflamed intestinal tissues, ulcerative colitis, Crohn's
 CC disease or cancer. The products can also be used for detection and
 CC diagnosis.
 CC
 SQ Sequence 389 BP; 105 A; 105 C; 102 G; 76 T; 1 other;
 XX
 XX
 QY Query Match 100.0%; Score 180; DB 19; Length 389;
 Db Best Local Similarity 100.0%; Pred. No. 1.2e-51;
 Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GGAAGGTCCTCTCTCAAGAAATGCCATCTGTGAACACATGTAGAGTCTCCAACTGT 60
 Db 112 GGAAGGTCCTCTCTCAAGAAATGCCATCTGTGAACACATGTAGAGTCTCCAACTGT 171
 QY 61 TCCAGATGTCACACCTGTGTCTGGCGCAGTGTGGCTCAGATATACGAATGCGCAG 120
 Db 172 TCCAGATGTCACACCTGTGTCTGGCGCAGTGTGGCTCAGATATACGAATGCGCAG 231
 QY 121 CTCTGCTTGGCCCGGATTAACCAACAGAGACATCCAGATCATGAAGATGCGCAATGC 180
 Db 232 CTCTGCTTGGCCCGGATTAACCAACAGAGACATCCAGATCATGAAGATGCGCAATGC 291
 XX
 XX
 RESULT 4
 ABV97573
 ID ABV97573 standard; cDNA; 394 BP.
 AC ABV97573;
 XX
 XX
 DT 14-JAN-2003 (first entry)
 DE Human pancreatic cancer expressed cDNA SEQ ID NO 2981.
 DE Human pancreatic cancer expressed cDNA SEQ ID NO 2981.
 KW Human; pancreas; cancer; gene therapy; vaccine; immunostimulant;
 KW cytosolic; tumour; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200260317-A2.
 PD 08-AUG-2002.
 XX
 PD 30-JAN-2002; 2002WO-US02781.
 PF 30-JAN-2002; 2002WO-US02781.
 XX
 PR 30-JAN-2001; 2001US-265305P.
 PR 31-JAN-2001; 2001US-265682P.
 PR 09-FEB-2001; 2001US-265682P.
 PR 21-MAR-2001; 2001US-278511P.
 PR 28-APR-2001; 2001US-287112P.
 PR 16-MAY-2001; 2001US-291631P.
 PR 12-JUL-2001; 2001US-305484P.
 PR 20-AUG-2001; 2001US-313999P.
 PR 27-NOV-2001; 2001US-333626P.
 XX
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Benson DR, Kalos MD, Lodes MJ, Bersing DH, Hepler WT, Jiang Y;
 XX WPI: 2002-627435/67.

XX New isolated polynucleotide and pancreatic tumor polypeptides, useful
 PT for diagnosing, preventing and/or treating cancer, particularly
 PT pancreatic cancer
 XX
 PS Claim 1; SEQ ID NO 2981; 300pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated polynucleotide (I) comprising: (a)
 CC any of a group of over 4000 nucleotide sequences (ABV94678-ABV99145);
 CC (b) complements of (a); (c) sequences consisting of at least 20
 CC contiguous residues of (a); (d) sequences that hybridize to (a), under
 CC moderately stringent conditions; (e) sequences having at least 75% or 90%
 CC identity to (a); or (f) degenerate variants of (a). Polypeptides
 CC (ABP68596-ABP68637) encoded by (i) and oligonucleotide can be used to
 CC detect cancer in a patient and compositions comprising polypeptides,
 CC polynucleotides, antibodies, fusion proteins, T cell populations and
 CC antigen presenting cells expressing the polypeptide are useful in
 CC treating pancreatic cancer and stimulating an immune response. The
 CC polynucleotides can be used as probes or primers for nucleic acid
 CC hybridisation, in the design and preparation of ribozyme molecules for
 CC inhibiting expression of the tumour polypeptides and proteins in the
 CC tumour cells, in vaccines and for gene therapy.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 394 BP; 115 A; 102 C; 100 G; 73 T; 4 other;
 XX
 XX
 QY Query Match 100.0%; Score 180; DB 24; Length 394;
 Db Best Local Similarity 100.0%; Pred. No. 1.2e-51;
 Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GGAAGGTCCTCTCTCAAGAAATGCCATCTGTGAACACATGTAGAGTCTCCAACTGT 60
 Db 77 GGAAGGTCCTCTCTCAAGAAATGCCATCTGTGAACACATGTAGAGTCTCCAACTGT 136
 QY 61 TCCAGATGTCACACCTGTGTCTGGCGCAGTGTGGCTCAGATATACGAATGCGCAG 120
 Db 137 TCCAGATGTCACACCTGTGTCTGGCGCAGTGTGGCTCAGATATACGAATGCGCAG 196
 QY 121 CTCTGCTTGGCCCGGATTAACCAACAGAGACATCCAGATCATGAAGATGCGCAATGC 180
 Db 197 CTCTGCTTGGCCCGGATTAACCAACAGAGACATCCAGATCATGAAGATGCGCAATGC 256
 XX
 XX
 RESULT 5
 AAT64673
 ID AAT64673 standard; cDNA to mRNA; 398 BP.
 AC AAT64673;
 XX
 XX
 DT 22-DEC-1997 (first entry)
 DE Human PEC-60-type protein encoding cDNA.
 DE Human PEC-60-type protein encoding cDNA.
 KW Human; PEC-60-type protein; digestive disorder; immune disorder;
 KW nervous disorder; gastrointestinal tract hormone; gastric tissue;
 KW antigen; antibody; ds.
 XX
 OS Homo sapiens.
 XX
 PN
 PD
 XX
 FH Key Location/Qualifiers
 FT CDS 43..303
 FT FT /*tag= a
 FT sig_peptide /product= PEC-60-type_protein
 FT mat_peptide 43..120
 FT 121..300
 FT /*tag= b
 FT /*tag= c
 XX
 XX
 PN WO9715596-A1.
 PD 01-MAY-1997.

XX 22-OCT-1996; 96MO-JP03061.
XX
XX 27-OCT-1995; 95JP-0280272.
XX
XX (SAGA) SAGAMI CHEM RES CENTRE,
XX
XX Kamata K, Kato S, Sekine-S, Yamaguchi T;
XX
XX WPI: 1997-258958/23.
XX P-PSDB: AAMW15067.
XX

PT Human PEC-60-type protein - for treatment or diagnosis of digestive,
PT Immune or nervous disorders
PS Claim 5; Page 10; 17pp; Japanese.

CC The present sequence encodes a human PEC-60-type protein. The protein
CC is a gastrointestinal tract hormone secreted by the gastric tissues.
CC It is useful for treatment or diagnosis of digestive, immune or nervous
CC disorders, or as an antigen to raise antibodies against them. Large
SO amounts of the protein may be produced by using the corresponding cDNA.

Sequence 398 BP; 107 A; 110 C; 105 G; 76 T; 0 other;

Query Match 100.0%; Score 180; DB 18; Length 398;
Best Local Similarity 100.0%; Pred. No. 1.2e-51;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0

OY 1 GGAAAGCTCCCTTTCGAAGAATGCCCATCTGTGAACAGATGGTAGACTCCCAACCTGT 60
Db | |||||
121 GGAAAGCTCCCTTTCGAAGAATGCCCATCTGTGAACAGATGGTAGACTCCCAACCTGT 180
OY 61 TCCAGATGTCACACCTGGTGCGGCAGCTGATGGGCTCACATATAAGCATGATGCCAG 120
Db | |||||
181 TCACAGATGTCACACCTGGTGCGGCAGCTGATGGGCTCACATATAAGCATGATGCCAG 240
OY 121 CTCGCTGTGGCCCCGATTAACCAACAAGACAGATCAGATATGAAGATGGCAAATGC 180
Db | |||||
241 CTCGCTGTGGCCCCGATTAACCAACAAGACAGATCAGATATGAAGATGGCAAATGC 300

RESULT 6
AAZ77490
ID AAZ77490 standard; cDNA, 402 BP.
XX
XX AAZ77490;
XX
XX 10-APR-2000 (first entry)
DE Human ovarian tumor cDNA library derived EST fragment 41.
XX
XX Expressed sequence tag; EST; human; ovarian tumor; anticancer;
KW gene therapy; treatment; ss.
OS Homo sapiens.
XX
XX DE19817557-A1.
PN
PD 21-OCT-1999.
PE 09-APR-1998; 98DE-1017557.
PF 09-APR-1998; 98DE-1017557.
PR (META-) METAGEN GES GENOMFORSCHUNG MBH.
PA
PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
DR WPI: 1999-591920/51.
DR P-PSDB: AAY76598, AAY76599.
XT New nucleic acid sequences expressed in ovarian, and some other, cancer

Query Match	Best Local Similarity	100.0%	Score 180;	DB 20;	Length 402;
Matches 180;	Conservative	100.0%	Pred. No. 1.2e-51;	Mismatches 0;	Indels 0; Gaps 0
QY	1	GGAAGAGCTGCGCTTCTCAAGATGCGCCATCTGTGAACATGCTGAGCTCCAACTGT	60		
Db	124	GGAAGAGCTGCGCTTCTCAAGATGCGCCATCTGTGAACATGCTGAGCTCCAACTGT	183		
QY	61	TCCAGATGTCACACCTGCTGCGGCACTGATGGGCTACATATACGAATGATGCCAG	120		
Db	184	TCCAGATGTCACACCTGCTGCGGCACTGATGGGCTACATATACGAATGATGCCAG	243		
QY	121	CTGCTGCTGGCGCCGGATATAAACCACCAACAGACATCCAGATCATTAAGATGGCAATGC	180		
Db	244	CTGCTGCTGGCGCCGGATATAAACCACCAACAGACATCCAGATCATTAAGATGGCAATGC	303		

RESULT 7
AAC98176
ID AAC98176 standard; cDNA: 433 BP.
AAC98176;
AAC98176;
09-MAR-2001 (first entry)

Human colon cancer antigen nucleotide sequence SEQ ID NO:186.

Human: colon cancer; colon cancer antigen; diagnosis; detection; identification; cytostatic; cardioactive; neuroprotective; vulnary; immunomodulatory; muscular; gynaecological; gastrointestinal; neoplastic; antineoplastic; antibacterial; gene therapy; wound; neural disorder; immune system disorder; muscular disorder; reproductive disorder; gastrointestinal disorder; renal disorder; infectious disease; cardiovascular disorder; ss.

Homo sapiens.
WO200055351-A1.
21-SEP-2000.
08-MAR-2000; 2000WO-US05883.
12-MAR-1999; 99US-0124270.
(HUMA-) HUMAN GENOME SCI INC.

Claim 3; Page 170; 310pp; German.

tissues and derived polypeptides, for treatment of ovarian cancer and identification of therapeutic agents -

[This invention describes novel nucleic acid (cDNA) sequences (A) which have antineoplastic activity and are highly expressed in ovarian tumor tissue (and some also in testis and breast cancer tissue). The products of the invention can be used for gene therapy. (A) are used (i) for recombinant expression of polypeptides (B) and (ii) to isolate complete genes. (B) are used (i) to identify agents suitable for treatment of ovarian cancer; (ii) directly for treating agents suitable for treatment of (including expression from gene therapy vectors) and (iii) for generation of specific antibodies. (A) are identified by assembling ESTs (expressed sequence tags) from a particular tissue type before comparison of the expression patterns. This allows a significantly longer fragment of the gene to be revealed, so should reduce the number of failures associated with the fact that ESTs from different libraries may represent different parts of the same unknown gene, distorting the estimated frequency of occurrence in a particular tissue. AA277450-277572 represent the human ovarian tumor cDNA library derived EST fragments described in the method of the invention and encode the protein fragments represented in AA276505-Y76638.

Sequence 402 BP; 108 A; 111 C; 107 G; 76 T; 0 other;

PI Rosen CA, Ruben SM;
 XX
 DR MPI: 2000-587534/55.
 DR P-PSDB: AAB53419.
 XX
 PT Colon cancer associated gene sequences, referred to as colon cancer
 PT antigens, useful for the treatment, prevention, and diagnosis of colon
 PT disorders such as colon cancer -
 XX
 PS Claim 1: Page 610; 2104pp; English.
 XX
 CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,
 CC called human colon cancer antigens, given in AAB53334 to AAB54006. The
 CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
 CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
 CC vulnerability, nephrologic, anti-infective and antibacterial activities, and
 CC can be used in gene therapy. The colon cancer antigen polynucleotides,
 CC proteins and antibodies to the proteins are useful for the prevention,
 CC treatment and diagnosis of colon disorders, such as colon cancer. The
 CC polynucleotides may be used in diagnostics and research, such as for
 CC chromosome identification, and as hybridisation probes. The proteins
 CC may also be used to prevent diseases such as neural disorders, immune
 CC system disorders, muscular disorders, reproductive disorders,
 CC gastrointestinal disorders, wounds, renal disorders, infectious
 CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
 CC AAB54007 represent sequences used in the exemplification of the present
 CC invention.
 CC
 XX Sequence 433 BP; 133 A; 113 C; 110 G; 77 T; 0 other;
 SQ
 Query Match 100.0%; Score 180; DB 21; Length 433;
 Best Local Similarity 100.0%; Pred. No. 1.2e-51;
 Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GGAAGCTCCCTTCTCAAGATGCCATCTGTGAACACATGTAGTCTCCAACTGT 60
 DB 137 GGAAGCTCCCTTCTCAAGATGCCATCTGTGAACACATGTAGTCTCCAACTGT 196
 QY 61 TCCGAGATGCCAAGCTGTGTGGGCACTGATGGGCTCATATATACGAATGATGCCAG 120
 DB 137 TCCGAGATGCCAAGCTGTGTGGGCACTGATGGGCTCATATATACGAATGATGCCAG 256
 QY 121 CTCTGCTTGGCCGGGATTAACCAACAGACATCCAGATCATCAAGATGGCAATGC 180
 DB 257 CTCTGCTTGGCCGGGATTAACCAACAGACATCCAGATCATCAAGATGGCAATGC 316

RESULT 8
 AAH34805
 ID AAH34805 standard; cDNA; 433 BP.
 XX
 AC AAH34805;
 XX
 DT 03-SEP-2001 (first entry)
 XX
 DE Human colon cancer antigen encoding cDNA SEQ ID NO:1887.
 XX
 KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KW colorectal carcinoma; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200122920-A2.
 XX
 PD 05-APR-2001.
 XX
 PF 28-SEP-2000; 2000WO-US26524.
 XX
 PR 29-SEP-1999; 99US-0157137.
 XX
 PR 03-NOV-1999; 99US-0163280.
 XX
 PA (HDMA-) HUMAN GENOME SCI INC.
 XX

PI Ruben SM, Barash SC, Birse CE, Rosen CA;
 XX
 DR MPI: 2001-235957/24.
 DR P-PSDB: AAG73400.
 XX
 PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 PT useful for preventing, diagnosing and/or treating colorectal cancers -
 XX
 PS Claim 1: Page 3403; 9803pp; English.
 XX
 CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing P.
 CC Inactive proteins or to supplement the patients own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated P,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC and AAG77789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.
 CC
 XX Sequence 433 BP; 133 A; 113 C; 110 G; 77 T; 0 other;
 SQ
 Query Match 100.0%; Score 180; DB 22; Length 433;
 Best Local Similarity 100.0%; Pred. No. 1.2e-51;
 Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GGAAGCTCCCTTCTCAAGATGCCATCTGTGAACACATGTAGTCTCCAACTGT 60
 DB 137 GGAAGCTCCCTTCTCAAGATGCCATCTGTGAACACATGTAGTCTCCAACTGT 196
 QY 61 TCCGAGATGCCAAGCTGTGTGGGCACTGATGGGCTCATATATACGAATGATGCCAG 120
 DB 137 TCCGAGATGCCAAGCTGTGTGGGCACTGATGGGCTCATATATACGAATGATGCCAG 256
 QY 121 CTCTGCTTGGCCGGGATTAACCAACAGACATCCAGATCATCAAGATGGCAATGC 180
 DB 257 CTCTGCTTGGCCGGGATTAACCAACAGACATCCAGATCATCAAGATGGCAATGC 316

RESULT 9
 ABO60075
 ID ABO60075 standard; cDNA; 468 BP.
 XX
 AC ABO60075;
 XX
 DT 02-AUG-2002 (first entry)
 XX
 DE Human colon cancer related nucleotide sequence SEQ ID NO:3770.
 XX
 KW Human; colon cancer; cancer; tissue profiling; forensic; mapping;
 KW genetic analysis; diagnostic; antisense therapy; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200229086-A2.
 XX
 PD 11-APR-2002.
 XX
 PF 02-OCT-2001; 2001WO-US30732.
 XX
 PR 02-OCT-2000; 2000US-237271P.
 XX
 PA (FARB) BAYER CORP.
 XX

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 23, 2003, 11:18:55; Search time 33.3218 Seconds
(without alignments)
2384.294 Million cell updates/sec

Title: US-09-941-309-1

Perfect score: 180

Sequence: 1 ggaagctcccttctcaag.....tcataagaatggaatgc 180

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapept 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

Issued Patents, NA: *

- 1: /cgn2_6/ptodata/1/ina/5A.COMB.seq: *
- 2: /cgn2_6/ptodata/1/ina/5B.COMB.seq: *
- 3: /cgn2_6/ptodata/1/ina/6A.COMB.seq: *
- 4: /cgn2_6/ptodata/1/ina/6B.COMB.seq: *
- 5: /cgn2_6/ptodata/1/ina/PCITUS.COMB.seq: *
- 6: /cgn2_6/ptodata/1/ina/backfile1.seq: *

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	180	100.0	180	US-09-065-019-1	Sequence 1, Appli
2	180	100.0	388	US-08-744-670-2	Sequence 2, Appli
3	180	100.0	388	US-09-149-933-2	Sequence 2, Appli
4	180	100.0	398	US-09-065-019-2	Sequence 2, Appli
5	39.8	22.1	458	US-08-880-829-21	Sequence 21, Appli
6	36.6	20.3	302	US-08-235-515A-26	Sequence 26, Appli
7	36.6	20.3	432	US-09-016-434-1250	Sequence 1250, Ap
8	35.4	19.7	420	US-09-221-298-56	Sequence 56, Appli
9	34.4	19.1	285	US-09-016-434-361	Sequence 361, App
10	32.2	17.9	265	US-09-221-298-55	Sequence 55, Appli
11	32	17.8	309	US-08-839-709-2	Sequence 2, Appli
12	32	17.8	309	US-09-204-859-2	Sequence 2, Appli
13	31.4	17.4	1129	US-09-383-586-10	Sequence 10, Appli
14	31.4	17.4	1206	US-09-312-283C-357	Sequence 357, App
15	31	17.2	126	US-08-586-676E-4	Sequence 4, Appli
16	31	17.2	129	US-08-586-676E-5	Sequence 5, Appli
17	31	17.2	138	US-08-586-676E-6	Sequence 6, Appli
18	30.6	17.0	3112	US-09-318-448-9	Sequence 9, Appli
19	29.4	16.3	180	US-08-744-670-2	Sequence 1, Appli
20	29.4	16.3	388	US-09-065-019-1	Sequence 2, Appli
21	29.4	16.3	388	US-08-744-670-2	Sequence 2, Appli
22	29.4	16.3	398	US-09-149-933-2	Sequence 2, Appli
23	29	16.1	276	US-09-065-019-2	Sequence 3, Appli
24	28.8	16.0	1230025	US-08-712-948-3	Sequence 3, Appli
25	28.4	15.8	463	US-09-198-452A-1	Sequence 1, Appli
26	28.4	15.8	463	US-09-439-313-362	Sequence 362, App
27	28.2	15.7	466	US-09-352-616A-362	Sequence 362, App
				US-09-141-027-5	Sequence 5, Appli

28	28.2	15.7	466	US-09-617-804-5	Sequence 5, Appli
29	28.2	15.7	771	US-08-972-008-3	Sequence 3, Appli
30	28.2	15.7	771	US-09-267-409-3	Sequence 3, Appli
31	28.2	15.7	2495	US-09-141-027-1	Sequence 1, Appli
32	28.2	15.7	2495	US-09-617-804-1	Sequence 1, Appli
33	28.2	15.7	2525	US-08-972-008-1	Sequence 1, Appli
34	28.2	15.7	2525	US-09-267-409-1	Sequence 1, Appli
35	27.4	15.2	2226	US-08-951-260A-1	Sequence 1, Appli
36	27.4	15.2	2226	US-09-430-626A-1	Sequence 1, Appli
37	27.4	15.2	2949	US-08-412-554A-3	Sequence 3, Appli
38	27.2	15.1	1529	US-08-821-278A-1	Sequence 1, Appli
39	27.2	15.1	1529	US-09-221-017B-506	Sequence 506, App
40	27	15.0	1988	US-08-712-948-8	Sequence 8, Appli
41	26.8	14.9	17000	US-09-679-299A-18	Sequence 18, Appli
42	26.6	14.8	2511	US-09-252-991A-9494	Sequence 9494, Ap
43	26.6	14.8	2583	US-09-252-991A-9541	Sequence 9541, Ap
44	26.6	14.8	4531	US-09-620-312D-893	Sequence 893, App
45	26.2	14.6	254	US-09-016-434-313	Sequence 313, App

ALIGNMENTS

RESULT 1
US-09-065-019-1
Sequence 1, Application US/09065019
Patent No. 6280968
GENERAL INFORMATION:
APPLICANT: Kato, Seishu
APPLICANT: Yamaguchi, Tomoko
APPLICANT: Sekine, Shingo
APPLICANT: Kamata, Kouju
TITLE OF INVENTION: HUMAN PEC-60-LIKE PROTEIN AND DNA ENCODING THIS PROTEIN
FILE REFERENCE: 6700PCT-US
CURRENT APPLICATION NUMBER: US/09/065, 019
CURRENT FILING DATE: 1998-04-17
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 180
TYPE: DNA
ORGANISM: Homo sapiens
US-09-065-019-1

Query Match 100.0%; Score 180; DB 3; Length 180;
Best Local Similarity 100.0%; Pred. No. 3.2e-56;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GGAAAGCTCCCTTTCTCAAGATGCCCATCTGTGTGAACATGTAGACTCTCCAACTGT	60
DB	1	GGAAAGCTCCCTTTCTCAAGATGCCCATCTGTGTGAACATGTAGACTCTCCAACTGT	60
QY	61	TCCAGATGTCACACCTGTGTGCGGACCTGATGAGGCTACATATACGAATGCGCAG	120
DB	61	TCCAGATGTCACACCTGTGTGCGGACCTGATGAGGCTACATATACGAATGCGCAG	120
QY	121	CTCTGCTTGCGCCGGATATAAACCAACAGACATCCAGATCATGAAGATGGCAATGC	180
DB	121	CTCTGCTTGCGCCGGATATAAACCAACAGACATCCAGATCATGAAGATGGCAATGC	180

RESULT 2
US-08-744-670-2
Sequence 2, Application US/08744670
Patent No. 5858710
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Goll, Surya K.
APPLICANT: Murry, Lynn E.
TITLE OF INVENTION: TUMOR-ASSOCIATED KAZAL INHIBITOR
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/744,670
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0155 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 388 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: SINTTUT01
CLONE: 1539065
US-08-744-670-2

Query Match 100.0%; Score 180; DB 2; Length 388;
Best Local Similarity 100.0%; Pred. No. 4.5e-56;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAAGCTCCCTTTCTCAAGATGCCCATCTGTGAACACATGTAGAGTCTCCAACTGT 60
DB 111 GGAAGCTCCCTTTCTCAAGATGCCCATCTGTGAACACATGTAGAGTCTCCAACTGT 170

QY 61 TCCAGATGTCCAACTGGTCTGGGACATGATGGGCTCAATATAGGAATGATGCCAG 120
DB 171 TCCAGATGTCCAACTGGTCTGGGACATGATGGGCTCAATATAGGAATGATGCCAG 230

QY 121 CTCTGCTTGGCCGGATTAACCAACAGACATCCAGTCAATGAAAGTGGCAATGC 180
DB 231 CTCTGCTTGGCCGGATTAACCAACAGACATCCAGTCAATGAAAGTGGCAATGC 290

RESULT 3
US-09-149-933-2
Sequence 2, Application US/09149933
Patent No. 5938699
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Goll, Surya K.
APPLICANT: Murry, Lynn E.
TITLE OF INVENTION: TUMOR-ASSOCIATED KAZAL INHIBITOR
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/149,933
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0155 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 388 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: SINTTUT01
CLONE: 1539065
US-09-149-933-2

Query Match 100.0%; Score 180; DB 2; Length 388;
Best Local Similarity 100.0%; Pred. No. 4.5e-56;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAAGCTCCCTTTCTCAAGATGCCCATCTGTGAACACATGTAGAGTCTCCAACTGT 60
DB 111 GGAAGCTCCCTTTCTCAAGATGCCCATCTGTGAACACATGTAGAGTCTCCAACTGT 170

QY 61 TCCAGATGTCCAACTGGTCTGGGACATGATGGGCTCAATATAGGAATGATGCCAG 120
DB 171 TCCAGATGTCCAACTGGTCTGGGACATGATGGGCTCAATATAGGAATGATGCCAG 230

QY 121 CTCTGCTTGGCCGGATTAACCAACAGACATCCAGTCAATGAAAGTGGCAATGC 180
DB 231 CTCTGCTTGGCCGGATTAACCAACAGACATCCAGTCAATGAAAGTGGCAATGC 290

RESULT 4
US-09-065-019-2
Sequence 2, Application US/09065019
Patent No. 6280968
GENERAL INFORMATION:
APPLICANT: Kato, Seishi
APPLICANT: Yamaguchi, Tomoko
APPLICANT: Sekine, Shingo
APPLICANT: Kamata, Kouju
TITLE OF INVENTION: HUMAN PEC-60-LIKE PROTEIN AND DNA ENCODING THIS PROTEIN
FILE REFERENCE: 6700PCT-US
CURRENT APPLICATION NUMBER: US/09/065,019
CURRENT FILING DATE: 1998-04-17
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 398
TYPE: DNA
ORGANISM: Homo sapiens
US-09-065-019-2

Query Match 100.0%; Score 180; DB 3; Length 398;
Best Local Similarity 100.0%; Pred. No. 4.5e-56;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAAGCTCCCTTTCTCAAGATGCCCATCTGTGAACACATGTAGAGTCTCCAACTGT 60
DB 121 GGAAGCTCCCTTTCTCAAGATGCCCATCTGTGAACACATGTAGAGTCTCCAACTGT 180

QY 61 TCCAGATGTCCAACTGGTCTGGGACATGATGGGCTCAATATAGGAATGATGCCAG 120
DB 181 TCCAGATGTCCAACTGGTCTGGGACATGATGGGCTCAATATAGGAATGATGCCAG 240

OY 121 CTCGCTGGCCGCGATTAACCAACAGACATCCAGATCAAGATGCAATGC 180
Db 241 CTCGCTGGCCGCGATTAACCAACAGACATCCAGATCAAGATGCAATGC 300

RESULT 5
US-08-880-829-21
Sequence 21, Application US/08880829
Patent No. 5923559

GENERAL INFORMATION:
APPLICANT: Collins, John
APPLICANT: Roettgen, Peter
TITLE OF INVENTION: A Collection of phagemids, A
TITLE OF INVENTION: Collection of Escherichia coli
TITLE OF INVENTION: Cells Carrying The Phagemids, A
TITLE OF INVENTION: Collection of Phagemid Particles
TITLE OF INVENTION: Produced From Said Collection
TITLE OF INVENTION: And Phagemid Particles
NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:
ADDRESSEE: Joseph T. Eisele
ADDRESSEE: Kane, Dalsimer, Sullivan, Kurucz,
ADDRESSEE: Levy, Eisele and Richard
STREET: 711 Third Avenue
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10017-4059
COMPUTER READABLE FORM:
MEDIUM TYPE: 3-1/2" DISKETTE
COMPUTER: IBM-XT COMPATIBLE
OPERATING SYSTEM: DOS 3.3:
SOFTWARE: WORDPERFECT 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/880, 829
FILING DATE: 23-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/458, 668
FILING DATE: 06/02/95
APPLICATION NUMBER: German EP 94 108 689.4
FILING DATE: 06/07/94
ATTORNEY/AGENT INFORMATION:
NAME: EISELE, JOSEPH T.
REGISTRATION NUMBER: 25,331
REFERENCE/DOCKET NUMBER: 2727-77
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 687-6000
TELEFAX: (212) 682-3485
TELEX: (212) 426767
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 458 base pairs
TYPE: nucleic acid
STRANDEDNESS: single strand
TOPOLOGY: linear
MOLECULE TYPE:
HYPOTHETICAL:
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM:
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
ORGANELLE:
IMMEDIATE SOURCE:
US-08-880-829-21

Query Match 22.1%; Score 39.8; DB 2; Length 458;
Best Local Similarity 57.7%; Fred. No. 6.2e-05;
Matches 71; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

OY 58 TGTTCACGATGTCACCTGCTGCGGCGACGTGAGGCTCACATATGCAATGAATGC 117
Db 220 TGCACCTAAGATCTACGACCCGCGTTGGCGTACCGACGCAACACTTACCCGACGAATGC 279
OY 118 CAGCTCTGCTGGCCGCGATTAACCAACAGACATCCAGATCAAGATGCAAA 177
Db 280 GTTCTGCTGCTGAACCAACCGTAACGTCAGACTCTATCTGATCCAGAAATCTGCTCG 339
OY 178 TGC 180
Db 340 TGC 342

RESULT 6
US-08-235-515A-26
Sequence 26, Application US/08235515A
Patent No. 5840518

GENERAL INFORMATION:
APPLICANT: Morishita, Hideaki
APPLICANT: Kanamori, Toshinori
TITLE OF INVENTION: DNA FRAGMENT, VECTOR CONTAINING THE DNA
TITLE OF INVENTION: FRAGMENT, TRANSFORMANT TRANSFORMED WITH THE VECTOR
TITLE OF INVENTION: PROCESS FOR PRODUCING PROTEIN USING THE VECTOR
NUMBER OF SEQUENCES: 65
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-3487

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/235, 515A
FILING DATE: 29-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1110-135P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-205-8050
TELEFAX: 703-205-8000
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 302 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: -
LOCATION: 1..302 /label= polynucleotide
OTHER INFORMATION: /note= "j-xi-y-z(pstl), insert in plasmid pm474,
OTHER INFORMATION: Figure 5"
FEATURE:
NAME/KEY: slg_peptide
LOCATION: 27..89
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 90..293
FEATURE:

```
OTHER INFORMATION: where n is a, c, g or t
FEATURE:
NAME/KEY: modified_base
```


; CURRENT APPLICATION NUMBER: US/60/485,450

;; CURRENT FILING DATE: 2003-07-09
;; NUMBER OF SEQ ID NOS: 47859
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 12159
;; LENGTH: 28753
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-60-485-450-12159

Query Match 65.2%; Score 117.4; DB 7; Length 28753;
Best Local Similarity 91.9%; Pred. No. 2.5e-31;
Matches 124; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 7 CTCCTTCACAGATGCCATCTGTGACACATGTTCCACCTGTTCCAG 66
Db 23973 CTCCTTCCTTCACAGCCCATCTGTGAACACATGTTCCACCTGTTCCAG 23914
OY 67 ATGTCCACCTGCTGCTGCGCAGTGTGATGGCTCATATACGATGAATGCGACCTGCG 126
Db 23913 ATGTCCACCTGCTGCTGCTGCGCAGTGTGATGGCTCATATACGATGAATGCGACCTGCG 23854
OY 127 TTGGCCCGATATAA 141
Db 23853 TTGGCCCGATATAA 23839

RESULT 3
US-60-493-007-3005
; Sequence 3005, Application US/60493007
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Higgs, Brandon
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; TITLE OF INVENTION: Primary rat hepatocyte toxicity modeling
; FILE REFERENCE: 44921-5113-01-US
; CURRENT APPLICATION NUMBER: US/60/493,007
; CURRENT FILING DATE: 2003-08-07
; PRIOR APPLICATION NUMBER: US 60/353,171
; PRIOR FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: US 60/363,534
; PRIOR FILING DATE: 2002-03-13
; PRIOR APPLICATION NUMBER: US 60/371,135
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: US 60/371,134
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: US 60/370,248
; PRIOR FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: US 60/371,150
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: US 60/371,413
; PRIOR FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: US 60/373,601
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/374,139
; PRIOR FILING DATE: 2002-04-22
; PRIOR APPLICATION NUMBER: US 60/394,253
; PRIOR FILING DATE: 2002-07-09
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 3518
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3005
; LENGTH: 7286
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. M64780
US-60-493-007-3005

Query Match 17.0%; Score 30.6; DB 7; Length 7286;
Best Local Similarity 73.6%; Pred. No. 1.6;

Matches 39; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
OY 70 TCCAACTGGTCTCCGCGCACTGATGGCTCATATACGATGAATGCCAGCT 122
Db 2683 TCTACCAAGGTCTCCGCGCACTGATGGCTCATATACGATGAATGCCAGCT 2735

RESULT 4
US-60-500-315-11222
; Sequence 11222, Application US/60500315
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING HUMAN G-PROTEIN COUPLED RECEPTOR PROTEINS, METHODS
; FILE REFERENCE: C1001484
; CURRENT APPLICATION NUMBER: US/60/500,315
; CURRENT FILING DATE: 2003-09-05
; NUMBER OF SEQ ID NOS: 69978
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11222
; LENGTH: 15859
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-500-315-11222

Query Match 16.8%; Score 30.2; DB 7; Length 15859;
Best Local Similarity 69.5%; Pred. No. 2.9;
Matches 41; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

OY 2 GAAAGCTCCCTTTCTCAAGATGCCATCTGTGTGACACATGTTAGATCTCCACCTGT 60
Db 4852 GAAAGCGCCCTCTCTCAAGATGCCATCTGTGTGACACATGTTAGATCTCCACCTGT 4910

RESULT 5
US-60-500-315-11231
; Sequence 11231, Application US/60500315
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING HUMAN G-PROTEIN COUPLED RECEPTOR PROTEINS, METHODS
; FILE REFERENCE: C1001484
; CURRENT APPLICATION NUMBER: US/60/500,315
; CURRENT FILING DATE: 2003-09-05
; NUMBER OF SEQ ID NOS: 69978
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11231
; LENGTH: 203938
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-500-315-11231

Query Match 16.8%; Score 30.2; DB 7; Length 203938;
Best Local Similarity 51.1%; Pred. No. 6.8;
Matches 71; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

OY 29 TCTGTGAACATGCTAGAGTCTTCAACCTGTTCCCGATGTCACACCTGCTGCGGCA 88
Db 176295 TCTGTGTGAACATGCTAGAGTCTTCAACCTGTTCCCGATGTCACACCTGCTGCGGCA 176354
OY 89 CTGATGGGCTCACATATACGATGAATGCGCAGCTGCTTGGCCGATATAAACCAAC 148
Db 176355 CAGCTTCTGTACAGTGTATTGTCAGAACCCACCTAAGCTGTGACAGGGAATAAAGCACTGAC 176414
OY 149 AGGACATCCAGATCATGAA 167
Db 176415 AGCAAAACAGAGCAGGTA 176433

RESULT 6
US-60-500-315-11538/c

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

OM nucleic - nucleic search, using sw model

9394.424 million cell updates/sec

Sequence: 1 gcagggcccgccagctcag.....taaaagatccagccaactg 398

Scoring table: IDENTITY_NUC

Total number of hits satisfying chosen parameters: 5777422

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

- 1: genembd1.*
- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_v1.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_v1.*
- 30: em_hcg_hum.*
- 31: em_hcg_inu.*
- 32: em_hcg_other.*
- 33: em_hcg_mus.*
- 34: em_hcg_pln.*
- 35: em_hcg_rod.*
- 36: em_hcg_ther.*
- 37: em_hcg_vrt.*
- 38: em_sy.*
- 39: em_hcg_hum.*
- 40: em_hcg_mus.*
- 41: em_hcg_ther.*

Pred. No. is the number of results predicted by chance to have a

Result No.	Score	Query Match	Length	DB	ID	Description
1	398	100.0	398	6	AR166138	AR166138 Sequence
2	398	100.0	398	6	E13091	E13091 Human cDNA
3	398	100.0	402	6	AX014851	AX014851 Sequence
4	388	97.5	388	6	AR028526	AR028526 Sequence
5	388	97.5	388	6	AR075817	AR075817 Sequence
6	388	97.5	388	6	BD007521	BD007521 Tumor-ass
7	367	92.2	386	9	AF048700	AF048700 Homo sapi
8	248.8	62.5	364	4	S4866	S4866 PRC-60-gast
9	248.8	62.5	364	4	SSEPC60	X67100 S.scrofa PB
10	211.2	53.1	515	10	MMMPG60	Y11505 M.musculus
11	180	45.2	180	6	AR166137	AR166137 Sequence
12	180	45.2	180	6	E13090	E13090 Human an cd
13	145.8	36.6	163396	9	AL161445	AL161445 Human DNA
14	145.8	36.6	176384	2	AC027812	AC027812 Homo sapi
15	145.8	36.6	183099	2	AC034144	AC034144 Homo sapi
16	117.4	29.5	188540	2	AC011108	AC011108 Homo sapi
17	81.2	20.4	192093	2	AC121305	AC121305 Rattus no
18	81.2	20.4	230493	10	AL837521	AL837521 Mouse DNA
19	81.2	20.4	242175	2	AC096463	AC096463 Rattus no
20	45.6	11.5	573	6	BD027885	BD027885 Sequence
21	45.6	11.5	621	6	BD027884	BD027884 Sequence
22	44.6	11.2	182	6	A11350	A11350 Artificial
23	44.6	11.2	182	6	A11352	A11352 Artificial
24	44.6	11.2	182	6	A24905	A24905 PEST master
25	44.6	11.2	182	12	SYNPSTIAA	M21832 Synthetic h
26	44.6	11.2	186	6	A24907	A24907 PEST master
27	42.2	10.6	368	6	AX332609	AX332609 Sequence
28	42.2	10.6	368	6	AX333348	AX333348 Sequence
29	42.2	10.6	368	6	AX411203	AX411203 Sequence
30	42.2	10.6	368	6	BD160949	BD160949 Preventiv
31	42.2	10.6	368	9	HSPST1	Y00705 Homo sapien
32	41.4	10.4	168	6	T08731	T08731 Sequence 19
33	41.4	10.4	482	6	E03479	E03479 cDNA encod1
34	41.4	10.4	5535	12	ASVPSKAN8	X84307 Artificial
35	41.2	10.4	369	10	MMPIP12	X06342 Mouse mRNA
36	40.6	10.2	341	6	E01574	E01574 cDNA sequen
37	40.6	10.2	362	9	BC025790	BC025790 Homo sapi
38	40.6	10.2	432	6	AR270687	AR270687 Sequence
39	40.6	10.2	432	6	H0MPSST1	M11949 Human pancr
40	40.6	10.2	608	9	AX014216	AX014216 Sequence
41	40.2	10.2	2200	6	AX655393	AX655393 Sequence
42	40.2	10.1	2228	10	AF094609	AF094609 Rattus no
43	39.6	9.9	179	6	E01725	E01725 Synthetic D
44	39.6	9.9	179	6	E02456	E02456 DNA encodin
45	39.6	9.9	204	12	SYNPSTIB	M31103 Synthetic h

ALIGNMENTS

RESULT	1
LOCUS	AR166138
DEFINITION	Sequence 2 from patent US 6280968.
ACCESSION	AR166138
VERSION	AR166138.1 GI:16241313
KEYWORDS	.
SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	Unclassified.
AUTHORS	1 (bases 1 to 398)
TITLE	Kato,S., Yamaguchi,T., Sekine,S. and Kamata,K.
JOURNAL	Human PEC-60-like protein and DNA encoding the same
FEATURES	Patent: US 6280968-A 2 28-AUG-2001;
	Location/Qualifiers

```

source          1. 398
BASE COUNT      107 a 110 c 105 g 76 t
ORIGIN
Query Match      100.0%; Score 398; DB 6; Length 398;
Best Local Similarity 100.0%; Pred. No.2.8e-99;
Matches 398; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGGCCCGCAGCGCTCAGGCTACACTATCCAGGATGAGCATGGCCGCTCCAGTGG 60
DB 1 GCAGGCCCGCAGCGCTCAGGCTACACTATCCAGGATGAGCATGGCCGCTCCAGTGG 60
QY 61 GTATATGCCCTGGCCCTTGGCTGCTGCTCTTGTGTGTGACAGGAGAGTGGCAGTGC 120
DB 61 GTATATGCCCTGGCCCTTGGCTGCTGCTCTTGTGTGTGACAGGAGAGTGGCAGTGC 120
QY 121 GGAAGCTCCCTTTCTCAAGAAATGCCATCTGTGAACACATGTTAGTCTCAACCTGT 180
DB 121 GGAAGCTCCCTTTCTCAAGAAATGCCATCTGTGAACACATGTTAGTCTCAACCTGT 180
QY 181 TCCAGATGTCCAACTGCTGTGCGGCACTGATGGCTCACATATACGAATGAATGCCAG 240
DB 181 TCCAGATGTCCAACTGCTGTGCGGCACTGATGGCTCACATATACGAATGAATGCCAG 240
QY 241 CTCTGCTTGGCCCGGATTAACCAACAGGACATCCAGATCATGTAAGATGGCAATGC 300
DB 241 CTCTGCTTGGCCCGGATTAACCAACAGGACATCCAGATCATGTAAGATGGCAATGC 300
QY 301 TGAATCCACAGAGGACCTCAACGCTCAACGCTCAAGTGTGAGAGACAGTGGTGGCATGG 360
DB 301 TGAATCCACAGAGGACCTCAACGCTCAACGCTCAAGTGTGAGAGACAGTGGTGGCATGG 360
QY 361 AGAGGATATGACATGAATAAAGATCCAGCCCACTG 398
DB 361 AGAGGATATGACATGAATAAAGATCCAGCCCACTG 398

RESULT 2
LOCUS          E13091          398 bp  DNA  linear  PAT 27-APR-1998
DEFINITION     Human cDNA encoding a PEC-60-like protein.
ACCESSION      E13091
VERSION        E13091.1 GI:3251903
KEYWORDS       JP 1997124698-A/2.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 398)
Kato,M., Yamaguchi,T., Sekine,S. and Kamata,M.
HUMAN PEC-60-LIKE PROTEIN AND DNA CAPABLE OF CODING THE SAME
Patent: JP 1997124698-A 2 13-MAY-1997;
SAGAMI CHEM RES CENTER
OS Homo sapiens (human)
PN JP 1997124698-A/2
PD 13-MAY-1997
PE 27-OCT-1995 JP 1995280272
PI KATO MASASHI, YAMAGUCHI TOMOKO, SEKINE SHINGO, PI KAMATA
MITSUHIKA
PC C07K14/575,C07H21/04,C12N15/09//A61K38/22,A61K38/22,A61K38/22,
PC C12P21/02,
PC (C12P21/02,C12R1:19);
CC strandedness: Double;
CC topology: Linear;
FH Key Location/Qualifiers
FT source 1. 398
FT /organism='Homo sapiens'
FT /tissue_type='stomach cancer' FT CDS
FT 43. .303 /product='PEC-60-like protein precursor' FT
FT sig_peptide 43. .120

```

```

FT mat_peptide 121. .300
FT /product='PEC-60-like protein'.
FEATURES
source          1. 398
Location/Qualifiers
/mol_type='genomic DNA'
/db_xref='taxon:9606'
BASE COUNT      107 a 110 c 105 g 76 t
ORIGIN
Query Match      100.0%; Score 398; DB 6; Length 398;
Best Local Similarity 100.0%; Pred. No.2.8e-99;
Matches 398; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGGCCCGCAGCGCTCAGGCTACACTATCCAGGATGAGCATGGCCGCTCCAGTGG 60
DB 1 GCAGGCCCGCAGCGCTCAGGCTACACTATCCAGGATGAGCATGGCCGCTCCAGTGG 60
QY 61 GTATATGCCCTGGCCCTTGGCTGCTGCTCTTGTGTGTGACAGGAGAGTGGCAGTGC 120
DB 61 GTATATGCCCTGGCCCTTGGCTGCTGCTCTTGTGTGTGACAGGAGAGTGGCAGTGC 120
QY 121 GGAAGCTCCCTTTCTCAAGAAATGCCATCTGTGAACACATGTTAGTCTCAACCTGT 180
DB 121 GGAAGCTCCCTTTCTCAAGAAATGCCATCTGTGAACACATGTTAGTCTCAACCTGT 180
QY 181 TCCAGATGTCCAACTGCTGTGCGGCACTGATGGCTCACATATACGAATGAATGCCAG 240
DB 181 TCCAGATGTCCAACTGCTGTGCGGCACTGATGGCTCACATATACGAATGAATGCCAG 240
QY 241 CTCTGCTTGGCCCGGATTAACCAACAGGACATCCAGATCATGTAAGATGGCAATGC 300
DB 241 CTCTGCTTGGCCCGGATTAACCAACAGGACATCCAGATCATGTAAGATGGCAATGC 300
QY 301 TGAATCCACAGAGGACCTCAACGCTCAACGCTCAAGTGTGAGAGACAGTGGTGGCATGG 360
DB 301 TGAATCCACAGAGGACCTCAACGCTCAACGCTCAAGTGTGAGAGACAGTGGTGGCATGG 360
QY 361 AGAGGATATGACATGAATAAAGATCCAGCCCACTG 398
DB 361 AGAGGATATGACATGAATAAAGATCCAGCCCACTG 398

RESULT 3
LOCUS          AX014851          402 bp  DNA  linear  PAT 07-SEP-2000
DEFINITION     Sequence 41 from Patent WO9953040.
ACCESSION      AX014851
VERSION        AX014851.1 GI:10041118
KEYWORDS       Homo sapiens (human)
SOURCE         Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1
Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
Pilarczyk,C.
Human nucleic acid sequences from ovarian tumour tissue
Patent: WO 9953040-A 41 21-OCT-1999;
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILARSKY CHRISTIAN (DE)
Location/Qualifiers
FT source 1. 402
FT /organism='Homo sapiens'
FT /mol_type='genomic DNA'
FT /db_xref='taxon:9606'
BASE COUNT      108 a 111 c 107 g 76 t
ORIGIN
Query Match      100.0%; Score 398; DB 6; Length 402;
Best Local Similarity 100.0%; Pred. No.2.8e-99;
Matches 398; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 GCAGCCCCCAGCAGCTAGGCTACACTATCCAGGATGACATGCGCCCTCCGCACTG 60
DB 4 GCAGCCCCCAGCAGCTAGGCTAGCTACACTATCCAGGATGACATGCGCCCTCCGCACTG 63
QY 61 GTATTCGCCCCCTGGCTGGCTGCCCCCTCTCTTGTGTGTGACAGGAAAGTCCAGTGGCAGCA 120
DB 64 GTATTCGCCCCCTGGCTGGCTGCCCCCTCTCTTGTGTGTGACAGGAAAGTCCAGTGGCAGCA 123
QY 121 GGAAGCTCCCTTTCTCAGAAATGCCATCTGTGAAACACATGTAGAGTCTCCAACTGT 180
DB 124 GGAAGCTCCCTTTCTCAGAAATGCCATCTGTGAAACACATGTAGAGTCTCCAACTGT 183
QY 181 TCCAGATGTCACACCTGCTGCTGGGACATGATGGGCTCACAATATAGAAATATGCCAG 240
DB 184 TCCAGATGTCACACCTGCTGCTGGGACATGATGGGCTCACAATATAGAAATATGCCAG 243
QY 241 CTCTGCTTGGCCCGGATTAACCAACAGAGACATCCAGATCATGAAGATGGCAATGC 300
DB 244 CTCTGCTTGGCCCGGATTAACCAACAGAGACATCCAGATCATGAAGATGGCAATGC 303
QY 301 TGATCCACAGAGACCTCAAGCCATGATGCTGAGAGAAACAGTGTGGGATG 360
DB 304 TGATCCACAGAGACCTCAAGCCATGATGCTGAGAGAAACAGTGTGGGATG 363
QY 361 AGAGATATGACATGAATTAAGATCCAGCCCAACTG 398
DB 364 AGAGATATGACATGAATTAAGATCCAGCCCAACTG 401

RESULT 4
AR028526
LOCUS AR028526 388 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 2 from patent US 5858710.
ACCESSION AR028526
VERSION AR028526.1 GI:5940499
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
SOURCE

Unknown.
Unclassified.
1 (bases 1 to 388)
Bandman, O., Goll, S.K. and Murry, L.E.
Tumor-associated kazal inhibitor-like polypeptides and encoding polynucleotides
Patent: US 5858710-A 2 12-JAN-1999;
Location/Qualifiers
1..388

BASE COUNT 105 a 105 c 102 g 76 t
ORIGIN

Query Match 97.5%; Score 388; DB 6; Length 388;
Best Local Similarity 100.0%; Pred. No. 1.6e-96;
Matches 388; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 GCCAGCTCAGGCTACACTATCCAGGATGACATGCGCGTCCGCACTGGTAATCGCCC 70
DB 1 GCCAGCTCAGGCTACACTATCCAGGATGACATGCGCGTCCGCACTGGTAATCGCCC 60
QY 71 TGGCCTTGGGCTGCCCTCTTGTGTGTGACAGGAAAGTCCAGTGGCAGCAAGAAAGCTCC 130
DB 61 TGGCCTTGGGCTGCCCTCTTGTGTGTGACAGGAAAGTCCAGTGGCAGCAAGAAAGCTCC 120
QY 131 CTTTCTCAAGATGCCATCTGTGTGAAACACATGTAGAGTCTCAACCTGTCCAGATGT 190
DB 121 CTTTCTCAAGATGCCATCTGTGTGAAACACATGTAGAGTCTCAACCTGTCCAGATGT 180
QY 191 CCAACTGTGCTGGGCACTGATGGGCTCACAATATAGAAATGCCAGCTGTGCTTGG 250
DB 181 CCAACTGTGCTGGGCACTGATGGGCTCACAATATAGAAATGCCAGCTGTGCTTGG 240
QY 251 CCGGATTAACCAACAGAGACATCCAGATCATGAAGATGGCAATGTGATCCACA 310
DB 111 CCGGATTAACCAACAGAGACATCCAGATCATGAAGATGGCAATGTGATCCACA 310

DB 241 CCGGATTAACCAACAGAGACATCCAGATCATGAAGATGGCAATGTGATCCACA 300
QY 311 GGAGCACTCAAGCCATGAAGTGTACCTGGAGAAACAGTGTGGCATGAGAGATATG 370
DB 301 GGAGCACTCAAGCCATGAAGTGTACCTGGAGAAACAGTGTGGCATGAGAGATATG 360
QY 371 ACATGAATTAAGATCCAGCCCAACTG 398
DB 361 ACATGAATTAAGATCCAGCCCAACTG 388

RESULT 5
AR075817
LOCUS AR075817 388 bp DNA linear PAT 30-AUG-2000
DEFINITION Sequence 2 from patent US 5958699.
ACCESSION AR075817
VERSION AR075817.1 GI:10002563
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
SOURCE

Unknown.
Unclassified.
1 (bases 1 to 388)
Bandman, O., Goll, S.K. and Murry, L.E.
Tumor-associated kazal inhibitor
Patent: US 5958699-A 2 28-SEP-1999;
Location/Qualifiers
1..388

BASE COUNT 105 a 105 c 102 g 76 t
ORIGIN

Query Match 97.5%; Score 388; DB 6; Length 388;
Best Local Similarity 100.0%; Pred. No. 1.6e-96;
Matches 388; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 GCCAGCTCAGGCTACACTATCCAGGATGACATGCGCGTCCGCACTGGTAATCGCCC 70
DB 1 GCCAGCTCAGGCTACACTATCCAGGATGACATGCGCGTCCGCACTGGTAATCGCCC 60
QY 71 TGGCCTTGGGCTGCCCTCTTGTGTGTGACAGGAAAGTCCAGTGGCAGCAAGAAAGCTCC 130
DB 61 TGGCCTTGGGCTGCCCTCTTGTGTGTGACAGGAAAGTCCAGTGGCAGCAAGAAAGCTCC 120
QY 131 CTTTCTCAAGATGCCATCTGTGTGAAACACATGTAGAGTCTCAACCTGTCCAGATGT 190
DB 121 CTTTCTCAAGATGCCATCTGTGTGAAACACATGTAGAGTCTCAACCTGTCCAGATGT 180
QY 191 CCAACTGTGCTGGGCACTGATGGGCTCACAATATAGAAATGCCAGCTGTGCTTGG 250
DB 181 CCAACTGTGCTGGGCACTGATGGGCTCACAATATAGAAATGCCAGCTGTGCTTGG 240
QY 251 CCGGATTAACCAACAGAGACATCCAGATCATGAAGATGGCAATGTGATCCACA 310
DB 241 CCGGATTAACCAACAGAGACATCCAGATCATGAAGATGGCAATGTGATCCACA 300
QY 311 GGAGCACTCAAGCCATGAAGTGTACCTGGAGAAACAGTGTGGCATGAGAGATATG 370
DB 301 GGAGCACTCAAGCCATGAAGTGTACCTGGAGAAACAGTGTGGCATGAGAGATATG 360
QY 371 ACATGAATTAAGATCCAGCCCAACTG 398
DB 361 ACATGAATTAAGATCCAGCCCAACTG 388

RESULT 6
BD007521
LOCUS BD007521 388 bp DNA linear PAT 31-JAN-2002
DEFINITION Tumor-associated kazal type inhibitor.
ACCESSION BD007521
VERSION BD007521.1 GI:18635894
KEYWORDS JP 2001503629-A/1.
SOURCE unidentified
ORGANISM unidentified

REFERENCE 1 (bases 1 to 388)
AUTHORS Bandman, O., Goli, S.K. and Murry, L.E.
TITLE Tumor-associated KAZAL type inhibitor
JOURNAL Patent: JP 2001503629-A 1 21-MAR-2001;
INCYTE PHARMACEUTICALS INC
COMMENT OS Unidentified
PN JP 2001503629-A/1
PD 21-MAR-2001
PR 05-NOV-1997 JP 1998521770
PR 06-NOV-1996 US 08/7744670
PI OLGA BANDMAN, SURYA K GOLI, LYNN E MURRY
PC C12N15/09, A61K38/55, A61K39/395, A61P1/00, A61P35/00, C07K14/81,
PC C07K16/38,
PC C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12Q1/68, G01N33/574, PC
C12N15/00, C12N5/00,
PC A61K37/64
CC
FH
FT
FEATURES
source
Location/Qualifiers
1..388
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
BASE COUNT 105 a 105 c 102 g 76 t
ORIGIN
Query Match 97.5%; Score 388; DB 6; Length 388;
Best Local Similarity 100.0%; Pred. No. 1.6e-96;
Matches 388; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 11 GCCAGCTCAGCTACACTATCCAGATCAGATGCGCCGTCCGCAAGTGGTAATCGCCC 70
DB 1 GCCAGCTCAGCTACACTATCCAGATCAGATGCGCCGTCCGCAAGTGGTAATCGCCC 60
QY 71 TGGCCTTGCTGCTCCCTCTCTTGTGTGACAGGGAAGTGCAGTGGCAGAGAAAGCTCC 130
DB 61 TGGCCTTGCTGCTCCCTCTCTTGTGTGACAGGGAAGTGCAGTGGCAGAGAAAGCTCC 120
QY 131 CTTTTCGAAGAATGCCCATATGTGACACATGTGTAAGTGTCCCACTGTCTCCAGATGT 190
DB 121 CTTTTCGAAGAATGCCCATATGTGTAACACATGTGTAAGTGTCCCACTGTCTCCAGATGT 180
QY 191 CCAACCTGCTGCTGGGCACTGATGGGCTCAGATCATATAGAAATGCAATGCACTGCTTGG 250
DB 181 CCAACCTGCTGCTGGGCACTGATGGGCTCAGATCATATAGAAATGCAATGCACTGCTTGG 240
QY 251 CCCGGATAAACCAACAGAGACATCCAGATCATGAAAGATGGCAATGCTGATCCACA 310
DB 241 CCCGGATAAACCAACAGAGACATCCAGATCATGAAAGATGGCAATGCTGATCCACA 300
QY 311 GGAGCAGCTCAAGCCATGAAGTGTCACTGAGAGACAGTGTGGGCAATGAGAGAGATATG 370
DB 301 GGAGCAGCTCAAGCCATGAAGTGTCACTGAGAGAGAGTGTGGGCAATGAGAGATATG 360
QY 371 ACATGAATTAAGATCCAGCCCACTG 398
DB 361 ACATGAATTAAGATCCAGCCCACTG 388
RESULT 7
AF048700 386 bp mRNA linear PRI 06-MAR-1998
LOCUS Homo sapiens gastrointestinal peptide (Pec-60) mRNA, complete cds.
DEFINITION
ACCESSION AF048700
VERSION AF048700.1 GI:2935439
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 386)
AUTHORS Hu, G.
TITLE Human gastrointestinal peptide mRNA, homolog of swine PEC-60
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 386)
AUTHORS Hu, G.
TITLE Direct Submission
JOURNAL Submitted (17-FEB-1998) Shanghai Institute of Cell Biology, 320
Yue-Yang Road, Shanghai 200031, China
COMMENT Location/Qualifiers
1..386
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1..386
/gene="PEC-60"
12..272
/gene="PEC-60"
/note="similar to Sus scrofa PEC-60"
/codon_start=1
/product="gastrointestinal peptide"
/protein_id="AAC05124.1"
/db_xref="GI:2935440"
/translation="MAVRQVYLAALALVYDREYVAAGKLPFSRMPICEHWESP
TCSQMSNVCGTGTGTYTNECQICLARIKTKDIDIMKDKC"
BASE COUNT 119 a 97 c 98 g 72 t
ORIGIN
Query Match 92.2%; Score 367; DB 9; Length 386;
Best Local Similarity 100.0%; Pred. No. 1.1e-90;
Matches 367; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 32 CCAGATCAGCATGCGCGTCCGCAAGTGGTAATCGCCCTTGAGCTTGCTGCTCCTTG 91
DB 1 CCAGATCAGCATGCGCGTCCGCAAGTGGTAATCGCCCTTGAGCTTGCTGCTCCTTG 60
QY 92 TTGTGACAGGAAGTCCAGTGGCAGCAGAGAAAGCTCCTTTCTCAAGATGCCATCT 151
DB 61 TTGTGACAGGAAGTCCAGTGGCAGCAGAGAAAGCTCCTTTCTCAAGATGCCATCT 120
QY 152 GTGAACACATGTAAGTCTCCAACTGTTTCCAGATGTCACACCTGTGTGGGCACTG 211
DB 121 GTGAACACATGTAAGTCTCCAACTGTTTCCAGATGTCACACCTGTGTGGGCACTG 180
QY 212 ATGGGCTCAGATATAGCAATGCAATGCGAGCTGCTGGCCGGATTAACCAACAGG 271
DB 181 ATGGGCTCAGATATAGCAATGCAATGCGAGCTGCTGGCCGGATTAACCAACAGG 240
QY 272 ACATCCAGATCATGAAAGATGGCAAAATGCTGATCCACAGAGCACCTCAAGCCATGAAG 331
DB 241 ACATCCAGATCATGAAAGATGGCAAAATGCTGATCCACAGAGCACCTCAAGCCATGAAG 300
QY 332 TGTCACTGAGAGAAAGTGTGGGCAATGAGAGAGATATGACATGAATTAAGATCCAGC 391
DB 301 TGTCACTGAGAGAAAGTGTGGGCAATGAGAGAGATATGACATGAATTAAGATCCAGC 360
QY 392 CCAACTG 398
DB 361 CCAACTG 367
RESULT 8
S46866 364 bp mRNA linear MAM 05-AUG-1999
LOCUS pEC-60-gastrointestinal peptide [swine, duodenum, mRNA, 364 nt].
DEFINITION
ACCESSION S46866
VERSION S46866.1 GI:257597
KEYWORDS
SOURCE Sus scrofa (pig)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

AUTHORS		TITLE		JOURNAL		REMARK	
Merris, M., Cintra, A., Solfrini, V., Entorf, P., Bortolotti, F., Morasuttil, D.G., Ostenson, C.G., Efendić, S., Agerberth, B., Mutt, V. et al.		Molecular cloning of PEC-60 and expression of its mRNA and peptide in the gastrointestinal tract and immune system		J. Biol. Chem. 267 (28), 19829-19832 (1992)		GenBank staff at the National Library of Medicine created this entry [NCBI gisbg 115614] from the original journal article. This sequence comes from Fig. 1.	
FEATURES		Location/Qualifiers					
source		1..364		/organism="Sus scrofa"			
gene		/mol_type="mRNA"		/db_xref="taxon:9823"			
CDS		1..364		/gene="PEC-60"			
		23..283		/gene="PEC-60"			
		/note="gastrointestinal peptide; This sequence comes from Fig. 1; Author includes translated amino acids, 5' of putative initiation codon"					
		/codon_start=1					
		/product="PEC-60"					
		/protein_id="AAB23691.2"					
		/db_xref="GI:5705936"					
		/translation="MAVRLMVLAALALITVDREVPVSAEKQYFSMPICEHMTESP DCSRIYDPVCIDGTGVYSESECKLCLAIENKQIDQIVKDGSC"					
BASE COUNT		91 a 93 c 108 g 72 t					
ORIGIN							
Query Match		62.5%; Score 248.8; DB 4; Length 364;					
Best Local Similarity		82.5%; Pred. No. 6.2e-58;					
Matches 297; Conservative		0; Mismatches 62; Indels 1; Gaps 1;					
OY	25	CACATATCCCAAGATCAGATGCGCGTCCGCAAGTGGTAATCGCCCTGGCTTGCGCC	84				
Db	5	CGCCAGCCCGAGATCAGATGCGCGTCCGCTGGGTGGTGGCCCTGGCTGGCTGCC	64				
OY	85	CTCCTTGTGGAGAGGAGATGCGAGTGGCAGAGAAAGCTCCCTTCTCAAGATG	144				
Db	65	CTCTTCATTTGGGACAGAGAGTCCAGTGTGGCAGAGAAAGCTTTCTCAAGATG	124				
OY	145	CCCATCTGTGAACATGTTAGTCTTCAACCTGTTCCCGATGTCCAACTGGTGTGC	204				
Db	125	CCCATCTGTGAACATGAGAGGATCCCGACAGTGTCCCGGATATAGACCACTGTCT	184				
OY	205	GGCAGTGGGCTCAGATATAGCATATGCAATGATGTCAGCTCTGCTGGGCCGATAAACC	264				
Db	185	GGCAGATGGGGTCAATATAGCAAGATGAAATGCAAGCTCTGCTGGCTCGATGAAACC	244				
OY	265	AAACAGGACATCCAGATCATGAAGATGGCAATGCTGATCCACAGAGACCTCAAGC	324				
Db	245	AAACAGGACATCCAGATCATGAAGATGGCAATGCTGATCCACAGAGACCTTAAAGC	304				
OY	325	CATGAAGTGTG-ACCTGGAGAACAGTGGTGGGCTGGAGAGATGATGATGAATAA	383				
Db	305	CATGAAGTGTGCGGCTGGAGAGACAGCTCTGGGTGTGGAGAAAGTGTGACATTAATAA	364				
RESULT 9							
SSPEC60		SSPEC60		364 bp		linear	
LOCUS		S.scrofa PEC-60 mRNA.					
ACCESSION		X67109					
VERSION		X67109.1 GI:2033					
KEYWORDS		pec-60 gene.					
SOURCE		Sus scrofa (p19)					
ORGANISM		Sus scrofa					
REFERENCE		Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.					
		(bases 1 to 364)					

[illegible]


```

PR      27-OCT-1995;       95JP-0280272.
XX
PA      (SAGA ) SAGAMI CHEM RES CENTRE.
XX
PI      Kamata K, Kato S, Sekine S, Yamaguchi T,
XX
DR      WPI; 1997-258958/23.
XX
P-PSDB; AAM15067.
XX
PT      Human PEC-60-type protein - for treatment or diagnosis of digestive,
XX
PS      immune or nervous disorders
XX
PC      Claim 5; Page 10; 17pp; Japanese.
XX
CC      The present sequence encodes a human PEC-60-type protein. The protein
CC      is a gastrointestinal tract hormone secreted by the gastric tissues..
CC      It is useful for treatment or diagnosis of digestive, immune or nervous
CC      disorders, or as an antigen to raise antibodies against them. Large
CC      amounts of the protein may be produced by using the corresponding cDNA.
XX
XX      Sequence 398 BP; 107 A; 110 C; 105 G; 76 T; 0 other;
XX
Query Match          100.0%; Score 398; DB 18; Length 398;
Best Local Similarity 100.0%; Pred. No. 1.2e-115;
Matches 398; Conservative 0; Mismatches 0; Indels 0; Gaps 0
XX
QY      1 GCAGGCCCCACGCGCTCAGGCTACACTATCCACAGATCAGCATGGCCGTCCGCAGTAGG 60
Db      1 GCAGGCCCCACGCGCTCAGGCTACACTATCCACAGATCAGCATGGCCGTCCGCAGTAGG 60
QY      61 GTATTGCGCCCTGGCCCTTGGCTGCCCTCTTTGTGTGACAGGGAAGTGCCACTGGCAGCA 120
Db      61 GTATTGCGCCCTGGCCCTTGGCTGCCCTCTTTGTGTGACAGGGAAGTGCCACTGGCAGCA 120
QY      121 GGAAAGCTCCCTTTCTCAAGAATGCCACTGTGTGAACACATGGTAGAGTCTCAACTGT 180
Db      121 GGAAAGCTCCCTTTCTCAAGAATGCCACTGTGTGAACACATGGTAGAGTCTCAACTGT 180
QY      121 GGAAAGCTCCCTTTCTCAAGAATGCCACTGTGTGAACACATGGTAGAGTCTCAACTGT 180
Db      121 GGAAAGCTCCCTTTCTCAAGAATGCCACTGTGTGAACACATGGTAGAGTCTCAACTGT 180
QY      181 TCCCAGATGTCCAACCTGTGTGGGCACTGATGGGCTTCACATATACGANTGATGCCAG 240
Db      181 TCCCAGATGTCCAACCTGTGTGGGCACTGATGGGCTTCACATATACGANTGATGCCAG 240
QY      241 CTCGCGCTGGCCCCGGATTAATAAACCAAAGCAGCATGCCAGTATGAAAGATGGCAAATGC 300
Db      241 CTCGCGCTGGCCCCGGATTAATAAACCAAAGCAGCATGCCAGTATGAAAGATGGCAAATGC 300
QY      301 CTCTGCTTTGGCCCCGGATTAATAAACCAAAGCAGCATGCCAGTATGAAAGATGGCAAATGC 300
Db      301 CTCTGCTTTGGCCCCGGATTAATAAACCAAAGCAGCATGCCAGTATGAAAGATGGCAAATGC 300
QY      301 TGATCCACAGAGACACCTCAAGCCATGAAGTGTGAGCTGGAGAACAGTGTGGGCATGG 360
Db      301 TGATCCACAGAGACACCTCAAGCCATGAAGTGTGAGCTGGAGAACAGTGTGGGCATGG 360
QY      361 AGAGGATATGACATGAATAAATAAGTCCAGGCCCACTG- 398
Db      361 AGAGGATATGACATGAATAAATAAGTCCAGGCCCACTG 398
XX
RESULT 2
AAZ77490 ID AAZ77490 standard; cDNA; 402 BP.
XX
AC.     AAZ77490;
XX
DT      10-APR-2000 (first entry)
XX
DE      Human ovarian tumor cDNA library derived EST fragment 41.
XX
RW      Expressed sequence tag; EST; human; ovarian tumor; anticancer;
KW      gene therapy; treatment; ss.
XX
OS      Homo sapiens.
XX
FN      DE19817557-A1.
XX
PD      21-OCT-1999.

```

[illegible]

AAC98176
ID AAC98176 standard; cDNA; 433 BP.
XX
AC AAC98176;
XX
DT 09-MAR-2001 (first entry)
XX
DE Human colon cancer antigen nucleotide sequence SEQ ID NO:186.
XX
KW Human: colon cancer; colon cancer antigen; diagnosis; detection;
KW identification; cytostatic; cardioactive; neuroprotective; vulnery;
KW immunomodulatory; muscular; gynaecological; gastrointestinal;
KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
KW neural disorder; immune system disorder; muscular disorder;
KW reproductive disorder; gastrointestinal disorder; renal disorder;
KW infectious disease; cardiovascular disorder; ss.
XX
OS Homo sapiens.
XX
PN MO20005351-A1.
XX
ID 21-SEP-2000.
XX
PE 08-MAR-2000; 2000WO-US05883.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI: 2000-587534/55.
DR P-PSDB: AAB53419.
XX
PT Colon cancer associated gene sequences, referred to as colon cancer
PT antigens, useful for the treatment, prevention, and diagnosis of colon
PT disorders such as colon cancer -
XX
XX
XX Claim 1; Page 610; 2104pp; English.
XX
PS AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC vulnery, nephrotropic, antiinfective and antibacterial activities, and
CC can be used in gene therapy. The colon cancer antigen polynucleotides,
CC proteins and antibodies to the proteins are useful for the prevention,
CC treatment and diagnosis of colon disorders, such as colon cancer. The
CC polynucleotides may be used in diagnostics and research, such as for
CC chromosome identification, and as hybridisation probes. The proteins
CC may also be used to prevent diseases such as neural disorders, immune
CC system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, wounds, renal disorders, infectious
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC AAB54007 represent sequences used in the exemplification of the present
CC invention.
XX
XX Sequence 433 BP; 133 A; 113 C; 110 G; 77 T; 0 other;
SQ
Query Match 99.2%; Score 395; DB 21; Length 433;
Best Local Similarity 100.0%; Pred. No. 1,1e-114;
Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 137 GGAAGCTCCCTTCTCAAGAAATGCCCATCTGTGACACATGTAAGTCTCCAACTGT 196
Qy 181 TCCAGATGTCACCACTGGTGTGGGACCTGATGGGCTACATATAGCAATAGCCAG 240
Db 197 TCCAGATGTCACCACTGGTGTGGGACCTGATGGGCTACATATAGCAATAGCCAG 256
Qy 241 CTCTGCTTGGCCGGATAAAAACCAACAGACATCCAGATCATGAATGCAATGC 300
Db 257 CTCTGCTTGGCCGGATAAAAACCAACAGACATCCAGATCATGAATGCAATAGC 316
Qy 301 TGATCCACAGACACCTCAAGCCATGAGTGTAGTGAGAACTGTGGGCATG 360
Db 317 TGATCCACAGACACCTCAAGCCATGAGTGTAGTGAGAACTGTGGGCATG 376
Qy 361 AGAGATATGACATGAATTAAGATCCAGCCCA 395
Db 377 AGAGATATGACATGAATTAAGATCCAGCCCA 411

RESULT 4
AAH34805
ID AAH34805 standard; cDNA; 433 BP.
XX
AC AAH34805;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen encoding cDNA SEQ ID NO:1887.
XX
KW Human: colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; ss.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
ID 05-APR-2001.
XX
PE 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI: 2001-235357/24.
DR P-PSDB: AAG75400.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
XX Claim 1; Page 3403; 9803pp; English.
XX
PS AAH32943 to AAH37195 and AAG773514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for

CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX Sequence 433 BP; 133 A; 113 C; 110 G; 77 T; 0 other;

Query Match 99.2%; Score 395; DB 22; Length 433;
Best Local Similarity 100.0%; Pred. No. 1.1e-114;
Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GCAGGCCCCAGCCAGCTGACATATCCAGATGAGATGCGCCGACGTGG 60
    |||||||
DB 17 GCAGGCCCCAGCCAGCTGACATATCCAGATGAGATGCGCCGACGTGG 76
    |||||||
QY 61 GATATGCGCCCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 120
    |||||||
DB 77 GATATGCGCCCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 136
    |||||||
QY 121 GGAAGCTCCCTTCTCAAGAAATGCCATCTGTAACACATGATAGATCTCAACCTGT 180
    |||||||
DB 137 GGAAGCTCCCTTCTCAAGAAATGCCATCTGTAACACATGATAGATCTCAACCTGT 196
    |||||||
QY 181 TCCCAAGATGTCACACCTGGTCTGGGCACTGATGGGCTCAATATACGAATGCGCAG 240
    |||||||
DB 197 TCCCAAGATGTCACACCTGGTCTGGGCACTGATGGGCTCAATATACGAATGCGCAG 256
    |||||||
QY 241 CTCTGCTGGCCGGATAAAAACCAACAGACATCCAGATCCATGAAAGATGGCAATGC 300
    |||||||
DB 257 CTCTGCTGGCCGGATAAAAACCAACAGACATCCAGATCCATGAAAGATGGCAATGC 316
    |||||||
QY 301 TGAATCCACAGAGACACCTCCAGCCATGAAATGTCAGCTGGAGAACAGTGGGGCATGG 360
    |||||||
DB 317 TGAATCCACAGAGACACCTCCAGCCATGAAATGTCAGCTGGAGAACAGTGGGGCATGG 376
    |||||||
QY 361 AGAGGATATGACATGAAATATAAAGATCCAGCCCA 395
    |||||||
DB 377 AGAGGATATGACATGAAATATAAAGATCCAGCCCA 411
    |||||||
```

RESULT 5
AAV38073
ID AAV38073 standard; cDNA; 389 BP.

AC AAV38073;

DT 14-SEP-1998 (first entry)

DE Human tumour-associated Kazal inhibitor encoding cDNA.

XX Human; tumour-associated Kazal inhibitor; TAK1; inflammation; cancer;
XX ulcerative colitis; Crohn's disease; inflammatory cytokine; elastase;
XX serine protease; glycoprotein hormone; trypsin; chymotrypsin; pepsin; ss.

OS Homo sapiens.

XX Key Location/Qualifiers
XX CDS 34..294
XX FT /*tag= a
XX FT /product= "tumour-associated Kazal inhibitor"

XX WO9820132-A1.

XX 14-MAY-1998.

XX 05-NOV-1997; 97WO-US20204.

XX 06-NOV-1996; 96US-0744670.

XX (INCY-) INCYTE PHARM INC.

XX Bandman O, Goli SK, Murry LE;

XX WPI; 1998-286948/25.

XX P-PDB; AAM62074.

PT New isolated tumour-associated Kazal inhibitor, TAK1 - used to
PT develop products for the diagnosis and treatment of e.g.
PT inflammation, ulcerative colitis, Crohn's disease or cancer
XX
PS Claim 5; Fig 1; 62pp; English.

CC The present sequence encodes human tumour-associated Kazal inhibitor
CC (TAK1). TAK1 is expressed in response to inflammatory cytokines and
CC functions in the inhibition of serine proteases or glycoprotein hormones
CC present in diseased tissues. The TAK1 can be used to inhibit the tissue
CC destruction associated with the production of excess proteases such as
CC trypsin, chymotrypsin, elastase, or pepsin. Products from the present
CC invention can be used in the treatment of e.g. proliferative cell
CC division in inflamed intestinal tissues, ulcerative colitis, Crohn's
CC disease or cancer. The products can also be used for detection and
CC diagnosis.

SO Sequence 389 BP; 105 A; 105 C; 102 G; 76 T; 1 other;

Query Match 97.5%; Score 388; DB 19; Length 389;
Best Local Similarity 100.0%; Pred. No. 1.7e-112;
Matches 388; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 11 GCCAGCTCAGGCTACACTATCCAGATCCAGATGCGCCGCAATGGGTATGCCCC 70
    |||||||
DB 2 GCCAGCTCAGGCTACACTATCCAGATCCAGATGCGCCGCAATGGGTATGCCCC 61
    |||||||
QY 71 TGGCCTTGGCTGGCCCTCTCTCTTGTGTGACAGGAAAGTGCAGTGGCAGCAAGCTCC 130
    |||||||
DB 62 TGGCCTTGGCTGGCCCTCTCTCTTGTGTGACAGGAAAGTGCAGTGGCAGCAAGCTCC 121
    |||||||
QY 131 CTTTCTCAGAAATGCCATCTGTGAACACATGATAGTGTCCAACTGTCCAGATGT 190
    |||||||
DB 122 CTTTCTCAGAAATGCCATCTGTGAACACATGATAGTGTCCAACTGTCCAGATGT 181
    |||||||
QY 191 CCAACCTGGTCTGGGACATGATGGGTCAATATGAAATGCACTGCTGTGG 250
    |||||||
DB 182 CCAACCTGGTCTGGGACATGATGGGTCAATATGAAATGCACTGCTGTGG 241
    |||||||
QY 251 CCGGATAAAAACCAACAGACATCCAGATCCAGAAAGTGGCAATGCTGATCCACA 310
    |||||||
DB 242 CCGGATAAAAACCAACAGACATCCAGATCCAGAAAGTGGCAATGCTGATCCACA 301
    |||||||
QY 311 GGAACACCTCAAGCCATGAAGTGTCACTGGAGACAGTGTGGCATGAGAGATATG 370
    |||||||
DB 302 GGAACACCTCAAGCCATGAAGTGTCACTGGAGACAGTGTGGCATGAGAGATATG 361
    |||||||
QY 371 ACATGAAATATAAAGATCCAGCCCACTG 398
    |||||||
DB 362 ACATGAAATATAAAGATCCAGCCCACTG 389
    |||||||
```

RESULT 6
ABV97573
ID ABV97573 standard; cDNA; 394 BP.

AC ABV97573;

DT 14-JAN-2003 (first entry)

DE Human pancreatic cancer expressed cDNA SEQ ID NO 2981.

XX Human; pancreas; cancer; gene therapy; vaccine; immunostimulant;
XX cytosolic; tumour; gene; ss.

OS Homo sapiens.

XX WO200260317-A2.

XX 08-AUG-2002.

XX 30-JAN-2002; 2002WO-US02781.

Accession	Source	Length	Score	DB	Length	Score	DB	
QY	240	GCTCGGCTGGTGGCCGGATTAACCAACAGGACATCC - AGATCAATGAAGATGCGCAAT	298					
Db	247	NCCTGCTTTGGCCGGATTAACCAACAGGACATCCAAATCATGAAAGATGGCAAT	306					
QY	299	GCTGATCCACAGGAGACCTCAACCCATGAAGTGTCACTGGAGAACAGTGTGGGCAT	358					
Db	307	GCTGATCCCAACAGGACCACTCAACCCATGAANTGTCACTTGGAGAACANTGTTGNGCA	366					
QY	359	GGAGA---GGATATGACATGGAATTAAGATCCAGCCCACT	397					
Db	367	TNGAGAGGATTTTGACATGGAATTAAGATCCAGCCCACT	408					
RESULT 8								
ABV86895/C	ABV86895 standard; cDNA; 323 BP.							
NC	ABV86895;							
XX	13-DEC-2002	(first entry)						
XX	Human colon cancer related cDNA SEQ ID NO 206.							
XX	Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine; gene;							
XX	ss.							
XX	Homo sapiens.							
XX	WO200258534-A2.							
XX	01-AUG-2002.							
XX	19-NOV-2001; 2001WO-US43704.							
XX	20-NOV-2000; 2000US-252222P.							
XX	06-FEB-2001; 2001US-267011P.							
XX	28-MAR-2001; 2001US-279670P.							
XX	10-JUL-2001; 2001US-304037P.							
XX	(CORI-) CORIXA CORP.							
XX	Stolk JA, Xu J, Chenault RA, Meagher MJ, Secretist H, King GE;							
XX	WPI; 2002-608400/65.							
XX	New isolated tumor colon polynucleotide and polypeptide, useful for the							
XX	diagnosis, prevention and/or treatment of cancer, in particular colon							
XX	cancer							
XX	Claim 1; SEQ ID NO 206; 266bp + Sequence Listing; English.							
XX	The invention relates to a human colon tumour expressed polynucleotide							
XX	(I) encoding a polypeptide (II, ABP67991-ABP67996) comprising: (I) any of							
XX	2600 fully defined nucleotide sequences (ABV8669-ABV89289); (II)							
XX	complements of (I); (III) at least 20 contiguous residues of (I); (IV)							
XX	sequences that hybridize to (I), under moderately stringent conditions;							
XX	(V) sequences having at least 75% or 90% identity to (I); or (VI)							
XX	degenerate variants of (I). The compositions and methods of the present							
XX	invention are useful for the diagnosis, prevention and/or treatment of							
XX	cancer, particularly colon cancer. (I) can be used in gene therapy and							
XX	(II) and (III) are useful in pharmaceutical compositions such as vaccines.							
XX	Note: The sequence data for this patent did not form part of the printed							
XX	specification, but was obtained in electronic format directly from WIPO							
XX	at ftp.wipo.int/pub/published_pct_sequences.							
XX	Sequence 323 BP; 66 A; 81 C; 93 G; 83 T; 0 other;							
XX	Query Match	80.8%;	Score	321.4;	DB	24;	Length	323;
XX	Best Local Similarity	99.7%;	Pred. No.	1,8e-91;				
XX	Matches	322;	Conservative	0;	Mismatches	1;	Indels	0;
XX	Gaps	0						

Db	323	GCCTCAGGCTCCTACCTATCCCGAGATCAGCATGSGCCCTCCGCCAGTGGATTAATCGCCCTGGC	264
OY	75	CTTGGCTGCGCCCTCTTGTGTGTGAGACAGGAAAGTGCAGTGGCAGAGAAAGCTCCCTTT	134
Db	263	CTTGGCTGCGCCCTCTTGTGTGTGAGACAGGAAAGTGCAGTGGCAGAGAAAGCTCCCTTT	204
OY	135	CTCAAGAAATGCCCATCTGTGTGAAACACATGCTAGAGTCTCCAACTCTTTCCCAAGTGTCAA	194
Db	203	CTCAAGAAATGCCCATCTGTGTGAAACACATGCTAGAGTCTCCAACTCTTTCCCAAGTGTCAA	144
OY	195	CTGTGCTGTGGGACATGATGGGCTCCACATATATGCAATGATGAGTCCAGCTCTGTGGGCGG	254
Db	143	CTGTGCTGTGGGACATGATGGGCTCCACATATATGCAATGATGAGTCCAGCTCTGTGGGCGG	84
OY	255	GATATAAAACCAACAGAGACATCCAGATCATGTAAGATGGAATGCTGATCCACAGAG	314
Db	83	GATATAAAACCAACAGAGACATCCAGATCATGTAAGATGGAATGCTGATCCACAGAG	24
OY	315	CACCTCAAGCCATGAAGTGTCAAG	337
Db	23	CACCTCAAGCCATGAAGTGTCAAG	1
RESULT 9			
ABV89019	ID	ABV89019	standard; cDNA: 323 BP.
XX	AC	ABV89019;	
XX	DT	13-DEC-2002	(first entry)
XX	DE	Human colon cancer related cDNA SEQ ID NO 2330.	
XX	DE	Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine; gene;	
XX	KW	ss.	
XX	OS	Homo sapiens.	
XX	PN	WO200258534-A2.	
XX	PD	01-AUG-2002.	
XX	PF	19-NOV-2001; 2001WO-0543704.	
XX	PR	20-NOV-2000; 2000US-252222P.	
XX	PR	06-FEB-2001; 2001US-267011P.	
XX	PR	28-MAR-2001; 2001US-279670P.	
XX	PR	10-JUL-2001; 2001US-304057P.	
XX	PA	(CORI-) CORIXA CORP.	
XX	XX	Stolk JA, Xu J, Chenault RA, Meagher MJ, Secrist H, King GE;	
XX	XX	WPI; 2002-608400/55.	
XX	PT	New isolated tumor colon polynucleotide and polypeptide, useful for the	
XX	PT	diagnosis, prevention and/or treatment of cancer, in particular colon	
XX	PT	cancer	
XX	PS	Claim 1; SEQ ID NO 2330; 266pp + Sequence Listing; English.	
XX	XX	The invention relates to a human colon tumour expressed polynucleotide	
XX	XX	(I) encoding a polypeptide (II, ABP67991-ABP67996) comprising: (i) any of	
XX	XX	2660 fully defined nucleotide sequences (ABV8669-ABV89289); (ii)	
XX	XX	complements of (i); (iii) at least 20 contiguous residues of (i); (iv)	
XX	XX	sequences that hybridize to (i), under moderately stringent conditions;	
XX	XX	(v) sequences having at least 75% or 90% identity to (i); or (vi)	
XX	XX	degenerate variants of (i). The compositions and methods of the present	
XX	XX	invention are useful for the diagnosis, prevention and/or treatment of	
XX	XX	cancer, particularly colon cancer. (i) can be used in gene therapy and	
XX	XX	(i) and (ii) are useful in pharmaceutical compositions such as vaccines.	
XX	XX	Note: The sequence data for this patent did not form part of the printed	

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 23, 2003, 11:16:55 : Search time 2358.39 seconds
(without alignments)
4101.602 Million cell updates/sec

Title: US-09-941-309-2
Perfect score: 398
Sequence: 1 gcacgccccccagcagctcag.....taaaagatccagcccaactg 398

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST: *
1: em_estba: *
2: em_esthm: *
3: em_estin: *
4: em_estcm: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_hic: *
9: gb_est1: *
10: gb_est2: *
11: gb_hic: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estom: *
17: em_gss_hum: *
18: em_gss_inv: *
19: em_gss_pln: *
20: em_gss_vit: *
21: em_gss_fun: *
22: em_gss_mam: *
23: em_gss_mus: *
24: em_gss_pro: *
25: em_gss_rtd: *
26: em_gss_phg: *
27: em_gss_vit: *
28: gb_gss1: *
29: gb_gss2: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	398	100.0	553	10	AM960685 EST372756
2	355	89.2	355	12	BM856263 K-EST0139
3	330.8	83.1	334	13	B0636767 he01b10.y
4	307.4	77.2	321	9	AA534438 nfb0b04.s

5	283.4	71.2	1115	10	BE543478
6	282	70.9	537	10	BG207464
7	227	57.0	232	7	AA296922
8	211.2	53.1	420	4	BY512489
9	207.2	52.1	220	4	AT732343
10	194	48.7	352	9	AA823442
11	190.6	47.9	364	13	BY089445
12	186.2	46.8	318	10	BE136647
13	185.2	46.5	330	9	AA692439
14	184.2	46.3	325	14	CB222030
15	184	46.2	395	9	AV066321
16	182.8	45.9	390	13	BY454425
17	182.2	45.8	361	9	AA591318
18	178.2	44.8	313	9	AA689871
19	177	44.5	210	12	BM856250
20	172.2	43.3	285	9	AA982819
21	170.2	42.8	316	9	AA821396
22	169.6	42.6	350	9	AA277286
23	167.8	42.2	433	9	AI009059
24	166	41.7	300	9	AA710853
25	163.4	41.1	370	9	AV066728
26	161.6	40.6	300	9	AV067091
27	161	40.5	293	12	BM430399
28	160.2	40.3	277	9	AV052389
29	153.2	38.5	277	9	AV067355
30	152.4	38.3	287	9	AV070840
31	150.6	37.8	369	9	AA733374
32	145.8	36.6	277	9	AV068882
33	145.8	36.6	582	9	AA502919
34	144.4	36.3	146	10	BF756812
35	144.2	36.2	542	9	AT732203
36	140.4	35.3	266	9	AV062346
37	137.8	34.6	242	9	AV066442
38	135.2	34.0	220	10	BE138001
39	135.2	34.0	229	9	AA592368
40	134.2	33.7	268	9	AA512376
41	130.8	32.9	214	10	BE199860
42	127.4	32.0	282	9	AV063285
43	127.4	32.0	286	9	AV064715
44	126.4	31.8	131	10	BG202409
45	119.4	30.0	301	9	AV068336

ALIGNMENTS

RESULT 1
LOCUS AM960685 553 bp mRNA linear EST 01-JUN-2000
DEFINITION EST372756 MAGE resequences, MAGF Homo sapiens CDNA, mRNA sequence.
ACCESSION AM960685
VERSION AM960685.1 GI:8150369
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 553)
Hegde, P., Ol, R., Abernathy, K., Dharp, S., Gaspard, R., Gay, C., Holt, I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and Quackenbush, J.
Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray

TITLE JOURNAL COMMENT
Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850-USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@igf.org
Plate: 149
Seq primer: Reverse.
Location/Qualifiers

FEATURES

```

source
1.553
/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/clone_lib="MAGE reserences, MAGE"
/note="Vector: BluescriptSkm"

BASE COUNT      156 a      153 c      137 g      107 t

Query Match      100.0%; Score 398; DB 10; Length 553;
Best Local Similarity 100.0%; Pred. No. 1.1e-101;
Matches 398; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGGCCAGCCAGCTACAGCTACCTATCCAGATCAGCATGGCCGTCGCCAGTGG 60
    |||||||
Db 3 GCAGGCCAGCCAGCTACAGCTACCTATCCAGATCAGCATGGCCGTCGCCAGTGG 62
    |||||||
QY 61 GTAATGCCCTGGCTTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 120
    |||||||
Db 63 GTAATGCCCTGGCTTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 122
    |||||||
QY 121 GGAAAGCTCCCTTTCTCAAGATGCCATCTGTGAACACATGGTAGAGTCTCAACCTGT 180
    |||||||
Db 123 GGAAAGCTCCCTTTCTCAAGATGCCATCTGTGAACACATGGTAGAGTCTCAACCTGT 182
    |||||||
QY 181 TCCCAATGTCACACTGCTGCTGGCAGTGTGATGGGCTCAGATATACGAATGATGCCAG 240
    |||||||
Db 183 TCCCAATGTCACACTGCTGCTGGCAGTGTGATGGGCTCAGATATACGAATGATGCCAG 242
    |||||||
QY 241 CTCTGCTGGCCCGGATTAACCAACAGAGCATCAGATCTGAAGATGGCAATGC 300
    |||||||
Db 243 CTCTGCTGGCCCGGATTAACCAACAGAGCATCAGATCTGAAGATGGCAATGC 302
    |||||||
QY 301 TGAATCCACAGAGACCTCAAGCCATGAAGTGTGAGTGTGAGAGACAGTGGTGGCATGG 360
    |||||||
Db 303 TGAATCCACAGAGACCTCAAGCCATGAAGTGTGAGTGTGAGAGACAGTGGTGGCATGG 362
    |||||||
QY 361 AGAGGATGATGACATGAATTAAGATCCAGCCCAACTG 398
    |||||||
Db 363 AGAGGATGATGACATGAATTAAGATCCAGCCCAACTG 400
    |||||||

RESULT 2
BM856263      355 bp      mRNA      linear      EST 06-MAR-2002
LOCUS      K-EST019933 S14K402 Homo sapiens cDNA clone S14K402-51-E10 5',
DEFINITION      mRNA sequence.
ACCESSION      BM856263
VERSION      BM856263.1 GI:19212662
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 355)
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
21C Frontier Korean EST Project 2001
Unpublished
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 51 row: E column: 10
High quality sequence stop: 355.
Location/Qualifiers
1.355
/organism="Homo sapiens"
/mol_type="mrna"

```

```

/db_xref="taxon:9606"
/clone="S14K402-51-E10"
/cell_line="K402"
/lab_host="Top10F"
/clone_lib="S14K402"
/note="Organ: Stomach; Vector: pTZ19Rpl; Site.1: EcoRI;
Site.2: NotI; The poly (A) RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tabacco acid pyrophosphatase (TAP). The decapped
interact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
clituarized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli top10F by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."

BASE COUNT      90 a      102 c      94 g      69 t

Query Match      89.2%; Score 355; DB 12; Length 355;
Best Local Similarity 100.0%; Pred. No. 1.3e-89;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 AGGCCCCAGCCAGCTACAGCTATCCAGATCAGCATGGCCGTCGCCAGTGGGT 62
    |||||||
Db 1 AGGCCCCAGCCAGCTACAGCTATCCAGATCAGCATGGCCGTCGCCAGTGGGT 60
    |||||||
QY 63 AATGCCCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 122
    |||||||
Db 61 AATGCCCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 120
    |||||||
QY 123 AAGGTCCTTTCTCAAGATGCCATCTGTGAACACATGGTGAAGTCTTCCAACCTGTTTC 182
    |||||||
Db 121 AAGGTCCTTTCTCAAGATGCCATCTGTGAACACATGGTGAAGTCTTCCAACCTGTTTC 180
    |||||||
QY 183 CCAGATGTCACACTGCTGCTGGCAGTGTGATGGGCTCAGATATACGAATGAATGCCAGCT 242
    |||||||
Db 181 CCAGATGTCACACTGCTGCTGGCAGTGTGATGGGCTCAGATATACGAATGAATGCCAGCT 240
    |||||||
QY 243 CTGCTTGGCCCGGATTAACCAACAGAGCATCAGATCATGAAGATGGCAATGCTG 302
    |||||||
Db 241 CTGCTTGGCCCGGATTAACCAACAGAGCATCAGATCATGAAGATGGCAATGCTG 300
    |||||||
QY 303 ATCCACAGAGAGACCTCAAGCCATGAAGTGTGAGTGTGAGAGACAGTGGGCA 357
    |||||||
Db 301 ATCCACAGAGAGACCTCAAGCCATGAAGTGTGAGTGTGAGAGACAGTGGGCA 355
    |||||||

RESULT 3
B0636767      334 bp      mRNA      linear      EST 15-JUL-2002
LOCUS      he01b10.y1 Human Retina cDNA (Un-normalized, unamplified): hd/he
DEFINITION      Homo sapiens cDNA clone he01b10 5', mRNA sequence.
ACCESSION      B0636767
VERSION      B0636767.1 GI:21761226
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 334)
Wistow,G., Bernstein,S.L., Wyatt,M.K., Ray,S., Behal,A., Touchman
J.W., Bouffard,G., Smith,D. and Peterson,K.
Expressed sequence tag analysis of human retina for the NEtBank
Project: Retbindin, an abundant, novel retinal cDNA and alternative
splicing of other retina-preferred gene transcripts
Mol. Vis. 8 (4), (2002) In press
Contact: Wistow G

```


Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: gramee@helix.nih.gov

Plate: 01 row: b column: 10
Seq primer: M13rpl reverse primer (ABI).

FEATURES

Location/Qualifiers

1..334
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="he01b10"
/tissue_type="Retina"
/dev_stage="Adult"
/lab_host="EMDH10B"
/clone_11b="Human Retina cDNA (un-normalized, unamplified)
); hd/he"

/note="Organ: Eye; Vector: pSPORT1; Neural retina tissue was dissected from two 80 year old donors with no observed eye disease. 100ug of total RNA was used for library construction. A directionally cloned cDNA library in the pSPORT1 vector (Life Technologies) was constructed at Bioserve Biotechnology (Laurel MD) essentially following the protocols of the Superscript Plasmid System full details of which are contained in the manufacturer's instruction manual (http://www.lifetech.com/). First strand synthesis was carried out using a Not I primer-adaptor [5'-pGACTAGTTCTGATCGGCGGCCGC(7)15-3'] . EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

BASE COUNT 94 a 86 c 88 g 66 t
ORIGIN

Query Match 83.1%; Score 330.8; DB 13; Length 334;
Best Local Similarity 99.4%; Pred. No. 8.7e-83;
Matches 332; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

59 GGGTAATCGCCCTGGCTGGCTGCTGCTGTTGTGTGAGAGGAAAGTCCAGTGGCAG 118
11 GGAACAATCGCCCTGGCTGGCTGCTGCTGTTGTGTGTGAGAGGAAAGTCCAGTGGCAG 60
119 CAGGAAGCTCCCTTTTCAGAGATGCCATCTGTGAACACATGTGTAGTCTCCACCT 178
61 CAGGAAGCTCCCTTTTCAGAGATGCCATCTGTGAACACATGTGTAGTCTCCACCT 120
179 GTTCCAGATGTCCACCTGCTGCGCACTGATGAGGCTCACAATATGCAATGAATGCC 238
121 GTTCCAGATGTCCACCTGCTGCGCACTGATGAGGCTCACAATATGCAATGAATGCC 180
239 AGCTCTGCTGGCCCGGATAAACAACAGACATCCAGATCATGAAGATGGCAAT 298
181 AGCTCTGCTGGCCCGGATAAACAACAGACATCCAGATCATGAAGATGGCAAT 240
299 GCTATCCACAGAGACCTCAAGCATGAAAGTCACTGGAGAACAGTGGGAGAT 358
241 GCTATCCACAGAGACCTCAAGCATGAAAGTCACTGGAGAACAGTGGGAGAT 300
359 GGAGAGATATGACATGAATGAATGAATGCCAGCC 392
301 GGAGAGATATGACATGAATGAATGAATGCCAGCC 334

RESULT 4
AA534438/c 321 bp mRNA linear EST 21-AUG-1997
LOCUS nfe8b04.s1 NCI CGAP Co3 Homo sapiens cDNA clone IMAGE:926191 3'
DEFINITION similar to SW:PF60_PIG P37109 PEPTIDE PEC-60 PRECURSOR. ; mRNA
sequence.
ACCESSION AA534438
VERSION AA534438.1 GI:2278691
KEYWORDS EST.

SOURCE

ORGANISM

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 321)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
Tumor Gene Index

Unpublished
Contact: Robert Strausberg, Ph.D.
Email: c9apb5-remail.nih.gov

Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILM at: www-bio.llnl.gov/bbrp/image/image.html

Insert length: 417 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amer sham.
Location/Qualifiers
1..321

FEATURES

Location/Qualifiers

1..321
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:926191"
/sex="pooled"
/tissue_type="colon"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from 12 pooled bulk tumor samples and primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization."

BASE COUNT 62 a 83 c 81 g 95 t
ORIGIN

Query Match 77.2%; Score 307.4; DB 9; Length 321;
Best Local Similarity 99.4%; Pred. No. 3.5e-76;
Matches 319; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

79 GCTGCCCTCTTGTGTGTGAGAGGAAAGTGGCAGTGGCAGAGAAAGCTCCCTTTCGA 138
321 GCTGCCCTCTTGTGTGTGAGAGGAAAGTGGCAGTGGCAGAGAAAGCTCCCTTTCGA 262
139 AGAATGCCATCTGTGAACACATGGTAGAGTCTCCAACTGTTCAGATGTCCAACTG 198
261 AGAATGCCATCTGTGAACACATGGTAGAGTCTCCAACTGTTCAGATGTCCAACTG 202
199 GTCTGGCGCACTGTGATGAGTCAATATGCAATG-AATGCCAGCTGCTGGCCGGAT 257
201 GTCTGGCGCACTGTGATGAGTCAATATGCAATG-AATGCCAGCTGCTGGCCGGAT 142
258 AAAAACCACAGAGACATCCAGATCATGAAGATGGCAATGCTGATCCACAGAGCAC 317
141 AAAAACCACAGAGACATCCAGATCATGAAGATGGCAATGCTGATCCACAGAGCAC 82
318 CTCAAGCCATGAAGTCACTGAGTGAAGACACTGTGGGATGAGAGAGATATGACATGA 377
81 CTCAAGCCATGAAGTCACTGAGTGAAGACACTGTGGGATGAGAGAGATATGACATGA 22
378 ATAAAGATCCAGCCCAACTG 398
21 ATAAAGATCCAGCCCAACTG 1

RESULT 5
BE543478

LOCUS BE543478 1115 bp mRNA linear EST 09-AUG-2000
 DEFINITION 601070792F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:3456839 5',
 mRNA sequence.
 ACCESSION BE543478
 VERSION BE543478.1 GI:9772123
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1115)
 NIH-MGC http://mhc.ncl.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 UNPUBLISHED
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
 http://image.llnl.gov
 Plate: LLAMB445 row: h column: 24
 High quality sequence stop: 385.
 Location/Qualifiers
 1..1115
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3456839"
 /tissue_type="cervical carcinoma cell line"
 /lab_host="DH10B"
 /clone_lib="NIH_MGC_12"
 /note="Organ: cervix; Vector: pCMV-SPORT6; Site:1; NotI;
 Site:2; SalI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 1.4 kb. Library prepared by Life
 Technologies."
 BASE COUNT 480 a 300 c 247 g 88 t
 ORIGIN
 Query Match 71.2%; Score 283.4; DB 10; Length 1115;
 Best Local Similarity 99.3%; Pred. No. 3.5e-69;
 Matches 235; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
 QY 103 GAAGTCCAGTGGCAGCAGAGAAAGCTCCCTTCTCAAGATGCCCATCTGTGAACATG 162
 Db 54 GAAGTCCAGTGGCAGCAGAGAAAGCTCCCTTCTCAAGATGCCCATCTGTGAACATG 113
 QY 163 GTAGAGTCTCCAACTGTTCCAGATGTCACACCTGCTGGCGCACTGATGGGCTCACA 222
 Db 114 GTAGAGTCTCCAACTGTTCCAGATGTCACACCTGCTGGCGCACTGATGGGCTCACA 173
 QY 223 TATACGAATGAATGCCAGCTCTGTCGCCCGGATATAAACCAGACGATCCAGATC 282
 Db 174 TATACGAATGAATGCCAGCTCTGTCGCCCGGATATAAACCAGACGATCCAGATC 233
 QY 283 ATGAAGATGGCAATAGCTGATCCCAAGAGAGACCTCAAGCCATGAAGTGTAGCTGGA 342
 Db 234 ATGAAGATGGCAATAGCTGATCCCAAGAGAGACCTCAAGCCATGAAGTGTAGCTGGA 293
 QY 343 GAACAGTGTGGG-CATGAGAGAGATATGACATGAATAAAGATCCAGCCCAACTG 398
 Db 294 GAACAGTGTGGGCTTGGAGAGAGATATGACATGAATAAAGATCCAGCCCAACTG 350
 RESULT 6
 BG207464 537 bp mRNA linear EST 21-APR-2001
 LOCUS RST26945 Atherys RAGE Library Homo sapiens cDNA, mRNA sequence.
 DEFINITION BG207464
 ACCESSION BG207464.1 GI:13729151
 KEYWORDS EST.

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 537)
 Harrington, J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R.,
 Cain, S., Leventhal, C., Thornton, M., Ramchandran, R., Whittington, J.,
 Lerner, L., Costanzo, D., McElligott, K., Booser, S., Mays, R., Smith,
 E., Veloso, N., Kika, A., Hess, J., Cochren, K., Lo, R., Offenbacher,
 J., Danzig, J. and Ducar, M.
 TITLE Creation of genome-wide protein expression libraries using random
 activation of gene expression
 Nat. Biotechnol. 19 (5), 440-445 (2001)
 JOURNAL 21227151
 MEDLINE 11329013
 PUBMED
 CONTACT: Scott J. Cain
 Atherys, Inc.
 3201 Carnegie Ave., Cleveland, OH 44115, USA
 Tel: 216 431 9900
 Fax: 216 361 9596
 Email: scain@atherys.com
 High quality sequence stop: 482.
 Location/Qualifiers
 1..537
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /cell_line="Hr1080"
 /clone_lib="Atherys RAGE Library"
 /note="See 'Creation of Genome-wide Protein Expression
 Libraries using Random Activation of Gene Expression',
 Nature Biotechnology, in press. Note that even though the
 cell type indicated is Hr1080, since a random activation
 method was used, these sequence tags are not necessarily
 expressed in Hr1080 under normal circumstances."
 BASE COUNT 150 a 133 c 148 g 106 t
 ORIGIN
 Query Match 70.9%; Score 282; DB 10; Length 537;
 Best Local Similarity 100.0%; Pred. No. 6.6e-69;
 Matches 282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 103 GAAGTCCAGTGGCAGCAGAGAAAGCTCCCTTCTCAAGATGCCCATCTGTGAACATG 162
 Db 252 GAAGTCCAGTGGCAGCAGAGAAAGCTCCCTTCTCAAGATGCCCATCTGTGAACATG 311
 QY 163 GTAGAGTCTCCAACTGTTCCAGATGTCACACCTGCTGGCGCACTGATGGGCTCACA 222
 Db 312 GTAGAGTCTCCAACTGTTCCAGATGTCACACCTGCTGGCGCACTGATGGGCTCACA 371
 QY 223 TATACGAATGAATGCCAGCTCTGTCGCCCGGATATAAACCAGACGATCCAGATC 282
 Db 372 TATACGAATGAATGCCAGCTCTGTCGCCCGGATATAAACCAGACGATCCAGATC 431
 QY 283 ATGAAGATGGCAATAGCTGATCCCAAGAGAGACCTCAAGCCATGAAGTGTAGCTGGA 342
 Db 432 ATGAAGATGGCAATAGCTGATCCCAAGAGAGACCTCAAGCCATGAAGTGTAGCTGGA 491
 QY 343 GAACAGTGTGGGATGAGAGATATGACATGAATAAAG 384
 Db 492 GAACAGTGTGGGATGAGAGATATGACATGAATAAAG 533
 RESULT 7
 AA296922
 LOCUS AA296922 232 bp mRNA linear EST 18-APR-1997
 DEFINITION EST112478 Retina II Homo sapiens cDNA 5' end similar to
 gastrointestinal peptide PEC-60, mRNA sequence.
 ACCESSION AA296922
 VERSION AA296922.1 GI:1949254
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

QY	1	GCAGGCCCCAGCAGCTCAGGCTACACTATCCAGAGATCAGCATGGCCCTCCGCACTGG	60
Db	1	GCAGGCCCCAGCAGCTCAGGCTACACTATCCAGAGATCAGCATGGCCCTCCGCACTGG	60
QY	61	GTATATGCCCTGGCCCTTGGCTGCCCTCTTGTGTGGACAGGGAATGCCAGTGGCAGCA	120
Db	61	GTATATGCCCTGGCCCTTGGCTGCCCTCTTGTGTGGACAGGGAATGCCAGTGGCAGCA	120
QY	121	GGAAACCTCCCTTTCACAGAAATGGCCATCTGTGAACACATGTGATGACTCTCCAAACCTGT	180
Db	121	GGAAACCTCCCTTTCACAGAAATGGCCATCTGTGAACACATGTGATGACTCTCCAAACCTGT	180
QY	181	TCCCAATGTCCAACCTGGTCTGCGGCACTGATGGGCTACATATATACGAATATGGCAG	240
Db	181	TCCCAATGTCCAACCTGGTCTGCGGCACTGATGGGCTACATATATACGAATATGGCAG	240
QY	241	CTCTGCTTGGCCCGGATAAAAACCAACGAGCATCCAGATCATGAAAGATGGCAATGC	300
Db	241	CTCTGCTTGGCCCGGATAAAAACCAACGAGCATCCAGATCATGAAAGATGGCAATATGC	300
QY	301	TGATGCCACAGAGAGACCTCTCAAGCCATGAAGTGTAGCTGGAGAAACACTGGTGGGCATGG	360
Db	301	TGATGCCACAGAGAGACCTCTCAAGCCATGAAGTGTAGCTGGAGAAACACTGGTGGGCATGG	360
QY	361	AGAGATATGACATGAATAAAGATGCCAGCCCAACTGG	398

Db 361 AGAGATATGACATGAATTAAGATCCAGCCCACTG 398

RESULT 2

US-08-744-670-2
Sequence 2, Application US/08744670
Patent No. 5858710

GENERAL INFORMATION:

APPLICANT: Bandman, Olga
APPLICANT: Goli, Surya K.
APPLICANT: Murry, Lynn E.
TITLE OF INVENTION: TUMOR-ASSOCIATED KAZAL INHIBITOR
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
City: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/744,670
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0155 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 388 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: SINTTUT01
CLONE: 1539065

US-08-744-670-2

Query Match 97.5%; Score 388; DB 2; Length 388;
Best Local Similarity 100.0%; Pred. No. 8.6e-118;

Matches 388; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 GCCAGCTCAGGCTACACTATCCAGATCGCATGCGCGTCCGCACTGGGTATGCCCC 70
Db 1 GCCAGCTCAGGCTACACTATCCAGATCGCATGCGCGTCCGCACTGGGTATGCCCC 60
QY 71 TGGCCTTGGCTGCGCTCCCTGTTGTGGACAGGGAAGTCCAGTGGGAGGAAAGCTCC 130
Db 61 TGGCCTTGGCTGCGCTCCCTGTTGTGGACAGGGAAGTCCAGTGGGAGGAAAGCTCC 120
QY 131 CTTTCTCAAGATGCGCATCTGTGAACACATGAGTGTCCCACTGTTCCAGATGT 190
Db 121 CTTTCTCAAGATGCGCATCTGTGAACACATGAGTGTCCCACTGTTCCAGATGT 180
QY 191 CCAACTGCTCTGCGGCACTGATGGGCTCACATATACGAATGAATGCCAGCTCTGCTTGG 250
Db 181 CCAACTGCTCTGCGGCACTGATGGGCTCACATATACGAATGAATGCCAGCTCTGCTTGG 240
QY 251 CCGGGAATTAAGATCCAGCCCACTG 398
Db 241 CCGGGAATTAAGATCCAGCCCACTG 388

QY 311 GGAGCACTTCAAGCCATGAAGTGTCACTGGAGAACAGTGGGCGCATGGAGAGATATG 370
Db 301 GGAGCACTTCAAGCCATGAAGTGTCACTGGAGAACAGTGGGCGCATGGAGAGATATG 360
QY 371 ACATGAATTAAGATCCAGCCCACTG 398
Db 361 ACATGAATTAAGATCCAGCCCACTG 388

RESULT 3

US-09-149-933-2

Sequence 2, Application US/09149933
Patent No. 5958699

GENERAL INFORMATION:

APPLICANT: Bandman, Olga
APPLICANT: Goli, Surya K.
APPLICANT: Murry, Lynn E.
TITLE OF INVENTION: TUMOR-ASSOCIATED KAZAL INHIBITOR
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
City: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/149,933
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0155 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 388 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: SINTTUT01
CLONE: 1539065

US-09-149-933-2

Query Match 97.5%; Score 388; DB 2; Length 388;
Best Local Similarity 100.0%; Pred. No. 8.6e-118;

Matches 388; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 GCCAGCTCAGGCTACACTATCCAGATCGCATGCGCGTCCGCACTGGGTATGCCCC 70
Db 1 GCCAGCTCAGGCTACACTATCCAGATCGCATGCGCGTCCGCACTGGGTATGCCCC 60
QY 71 TGGCCTTGGCTGCGCTCCCTGTTGTGGACAGGGAAGTCCAGTGGGAGGAAAGCTCC 130
Db 61 TGGCCTTGGCTGCGCTCCCTGTTGTGGACAGGGAAGTCCAGTGGGAGGAAAGCTCC 120
QY 131 CTTTCTCAAGATGCGCATCTGTGAACACATGAGTGTCCCACTGTTCCAGATGT 190
Db 121 CTTTCTCAAGATGCGCATCTGTGAACACATGAGTGTCCCACTGTTCCAGATGT 180
QY 191 CCAACTGCTCTGCGGCACTGATGGGCTCACATATACGAATGAATGCCAGCTCTGCTTGG 250
Db 181 CCAACTGCTCTGCGGCACTGATGGGCTCACATATACGAATGAATGCCAGCTCTGCTTGG 240

```

Db      181  CCAACTGCTGCGGCACTGATGGGCTCACATATACAAATGAATGCCAGCTCTCTTGG 240
Qy      251  CCCGATATAAAAACCAAAAGAGCATCCGATCATATAAAGATGGCAAAATGCTGATCCCA 310
Db      241  CCCGATATAAAAACCAAAAGAGCATCCGATCATATAAAGATGGCAAAATGCTGATCCCA 300
Qy      311  GGAGCAGCTCAAGCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 370
Db      301  GGAGCAGCTCAAGCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 360
Qy      371  ACATGAATATAAAGATCCAGCCCAACTG 398
Db      361  ACATGAATATAAAGATCCAGCCCAACTG 388

```

RESULT 4

US-09-065-019-1

Sequence 1, Application US/09065019

Patent No. 6280968

GENERAL INFORMATION:

APPLICANT: Kato, Seishi

APPLICANT: Yamaguchi, Tomoko

APPLICANT: Sekine, Shingo

APPLICANT: Kamata, Kouju

TITLE OF INVENTION: HUMAN PEC-60-LIKE PROTEIN AND DNA ENCODING THIS PROTEIN

FILE REFERENCE: 6700PCT-US

CURRENT APPLICATION NUMBER: US/09/065.019

CURRENT FILING DATE: 1998-04-17

NUMBER OF SEQ ID NOS: 3

SOFTWARE: Patentl Ver. 2.0

SEQ ID NO 1

LENGTH: 180

TYPE: DNA

ORGANISM: Homo sapiens

US-09-065-019-1

Query Match 45.2%; Score 180; DB 3; Length 180;

Best Local Similarity 100.0%; Pred. No. 1.1e-49;

Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy      121  GGAAGCTCCCTTCTCAGAGATGCCATCTGTGAACAGATGTAGATCTCCAACTGT 180
Db      1  GGAAGCTCCCTTCTCAGAGATGCCATCTGTGAACAGATGTAGATCTCCAACTGT 60
Qy      181  TCCAGATGTCCAACTGCTGCGGCACTGATGGCTCACATATACGAATGAATGCCAG 240
Db      61  TCCAGATGTCCAACTGCTGCGGCACTGATGGCTCACATATACGAATGAATGCCAG 120
Qy      241  CTCGCTTGCGCCGCGATATAAACCAGAGACATCCAGATCATGAAGAATGCCAAATGC 300
Db      121  CTCGCTTGCGCCGCGATATAAACCAGAGACATCCAGATCATGAAGAATGCCAAATGC 180

```

RESULT 5

US-08-880-829-21

Sequence 21, Application US/08880829

Patent No. 5925559

GENERAL INFORMATION:

APPLICANT: Collins, John

APPLICANT: Roettgen, Peter

TITLE OF INVENTION: A Collection of Phagmids, A

TITLE OF INVENTION: Collection of Escherichia Coli

TITLE OF INVENTION: Cells Carrying The Phagmids, A

TITLE OF INVENTION: Collection of Phagmid Particles

TITLE OF INVENTION: Produced From Said Collection

TITLE OF INVENTION: And Phagmid Particles

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: Joseph T. Eisele

ADDRESSEE: Kane, Dalsimer, Sullivan, Kurucz,

ADDRESSEE: Levy, Eisele and Richard

STREET: 711 Third Avenue

STREET: 711 Third Avenue

```

CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10017-4059
COMPUTER READABLE FORM:
MEDIUM TYPE: 3-1/2" DISKETTE
COMPUTER: IBM-XT COMPATIBLE
OPERATING SYSTEM: DOS 3.3
SOFTWARE: WORDPERFECT 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/880.829
FILING DATE: 23-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/458,668
FILING DATE: 06/02/95
APPLICATION NUMBER: German EP 94 108 689.4
FILING DATE: 06/07/94
ATTORNEY/AGENT INFORMATION:
NAME: EISELE, JOSEPH T.
REGISTRATION NUMBER: 25,331
REFERENCE/DOCKET NUMBER: 2727-77
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 687-6000
TELEFAX: (212) 682-3485
TELEX: (212) 426767
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 458 base pairs
TYPE: nucleic acid
STRANDEDNESS: single strand
TOPOLOGY: linear
MOLECULE TYPE:
HYPOTHETICAL:
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM:
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE:
US-08-880-829-21

```

Query Match

Best Local Similarity 10.4%; Score 41.4; DB 2; Length 458;

Matches 75; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

```

Qy      178  TGTTCAGATGTCCAACTGCTGCGGCACTGATGGGCTCACATATACGAATGATCC 237
Db      220  TGCATAGATGTCCAACTGCTGCGGCACTGATGGGCTCACATATACGAATGATCC 279
Qy      238  CAGCTGCTTGCGCCGCGATATAAACCAGAGACATCCAGATCATGAAGAATGCCAAA 297
Db      280  GTTCTGCTTGCGCCGCGATATAAACCAGAGACATCCAGATCATGAAGAATGCCAAA 339
Qy      298  TCGTGAATCCA 308
Db      340  TCGTGAATCCA 350

```

RESULT 6

US-09-016-434-1250

Sequence 1250, Application US/09016434

Patent No. 6500938

GENERAL INFORMATION:

APPLICANT: Janice Au-Young

APPLICANT: Jeffrey J. Seilhamer

```

Query Match      10.2% ; Score 40.6 ; DB 4 ; Length 432 ;
Best Local Similarity 56.3% ; Pred. No. 0.001 ;
Matches 76 ; Conservative 0 ; Mismatches 59 ; Indels 0 ; Gaps 0 ;

QY      178  TGTTCACAGTATGTCACACTGTGCTGTGGCAGCATGATGGGCTACATATATAGAAATATGC 237
           | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      231  TGACCCAGATATATATGACCTGTCTGTGGGACGTGATGGAATATCTTATCCCAATGAATGC 290
           | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      238  CAGCTCTGCTGTGGCCCCGATATAAACCAACAGACATCCAGATCTATGAAAGATGGCAAA 297
           | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      291  GTGTATATGTTTGTGAANAATCGGAACGCGACACATCTTATCTCATTTCAAAAATCTGGGCT 350
           | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      298  TGTGATCCCAACAG 312
           | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      351  TGCTGAGAACCAAG 365
           | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 7
: Sequence 26, Application US/08235515A
: Patent No. 5840518
: GENERAL INFORMATION:
:   APPLICANT: Morishita, Hideaki
:   APPLICANT: Kanamori, Toshinori
:   APPLICANT: No. 5840518uhara, Masahiro
:   TITLE OF INVENTION: DNA FRAGMENT, VECTOR CONTAINING THE DNA
:   TITLE OF INVENTION: FRAGMENT, TRANSFORMANT TRANSFORMED WITH THE VECTOR AND
:   NUMBER OF SEQUENCES: 65
:   CORRESPONDENCE ADDRESS:
:   ADDRESSEE: Birch, Stewart, Kolasch & Birch
:   STREET: P.O. Box 747

```

```

1 CITY: Falls Church
2 STATE: Virginia
3 COUNTRY: USA
4 ZIP: 22040-3487
5 COMPUTER READABLE FORM:
6 MEDIUM TYPE: Floppy disk
7 COMPUTER: IBM PC compatible
8 OPERATING SYSTEM: PC-DOS/MS-DOS
9 SOFTWARE: Patentin Release #1.0, Version #1.25
10 CURRENT APPLICATION DATA:
11 APPLICATION NUMBER: US/08/235,515A
12 FILING DATE: 29-APR-1994
13 CLASSIFICATION: 435
14 ATTORNEY/AGENT INFORMATION:
15 NAME: Murphy Jr., Gerald M.
16 REGISTRATION NUMBER: 28,977
17 REFERENCE/DOCKET NUMBER: 1110-135P
18 TELECOMMUNICATION INFORMATION:
19 TELEPHONE: 703-205-8000
20 TELEFAX: 703-205-8050
21 INFORMATION FOR SEQ ID NO: 26:
22 SEQUENCE CHARACTERISTICS:
23 LENGTH: 302 base pairs
24 TYPE: nucleic acid
25 STRANDEDNESS: double
26 TOPOLOGY: linear
27 MOLECULE TYPE: cDNA
28 FEATURE:
29 NAME/KEY: -
30 LOCATION: 1..302
31 OTHER INFORMATION: /label= polynucleotide
32 OTHER INFORMATION: /note= "j-xl-y-z(pstl), insert in plasmid pm474,
33 OTHER INFORMATION: Figure 5"
34 FEATURE:
35 NAME/KEY: sig_peptide
36 LOCATION: 27..89
37 FEATURE:
38 NAME/KEY: mat_peptide
39 LOCATION: 90..293
40 FEATURE:
41 NAME/KEY: CDS
42 LOCATION: 27..293
43 US-08-235-515A-26
44
45 Query Match          9.9%; Score 39.6; DB 2; Length 302;
46 Best Local Similarity 57.1%; Pred. No. 0.0018;
47 Matches 72; Conservative 0; Mismatches 54; Indels 0; Gaps 0
48
49 OY      178 TGTTCGCCAGATGTCACACCTGGTCTGGCGGCACTGATGGGCTCACAATAACGATGAATGC 237
50           ||| |||| |   ||| |||| |   ||| |||| |   ||| |||| |
51 Db       171 TGACACCAAGATATATGACCCTGCTCTGTGGACTGATGAAATCTTAATCCCATGATGC 230
52
53 OY      238 CAGCTCTGCTTGCGCCCGATATAAACCAACAGACGACATCCAATCATGAAGAATGGCAA 297
54           ||| |||| |   ||| |||| |   ||| |||| |   ||| |||| |
55 Db       231 GTGTTATCTTTGAANAATGGAAACGCCAGACATGCATCTCATTCAAAANAATCTGGGCCT 290
56
57 OY      298 TGCTGA 303
58           |||||
59 Db       291 TGCTGA 296
60
61 RESULT 8
62 US-09-221-298-56/c
63 Sequence 56, Application US/09221298
64 Patent No. 6284241
65 GENERAL INFORMATION:
66 APPLICANT: Xu, Jiangchun
67 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY AND DIAGNOSIS
68 FILE REFERENCE: 210121.471
69 CURRENT APPLICATION NUMBER: US/09/221,298
70 CURRENT FILING DATE: 1998-12-23
71 NUMBER OF SEQ ID NOS: 112

```